Identification and Characterization of Genetic Variants Conveying Risk to Develop Uterine Leiomyomata

Dissertation Advisor: Cynthia Casson Morton Author: Stacey Lynn Eggert

Identification and Characterization of Genetic Variants

Conveying Risk to Develop Uterine Leiomyomata

ABSTRACT

Uterine leiomyomata (UL), commonly known as fibroids, are a major public health problem given their extreme prevalence (>70%), severity of associated symptoms, and indication for hysterectomies in women of reproductive age. Familial aggregation and twin studies have provided evidence for a genetic component to predisposition to develop UL. To date, a small number of genes involved in UL biology, including *HMGA2*, have been discovered through cytogenetic studies of the tumors. *HMGA2* is involved in recurrent translocations in UL and a TC repeat polymorphism in the gene is associated with UL diagnosis. In this thesis, I investigate the possible role of the TC repeat in *HMGA2* expression. In 293T cells, the TC repeat number did not affect promoter activation, however, in the more relevant UL and myometrial cells, *HMGA2* promoter activation was severely impaired and a definitive conclusion could not be made. Genome-wide linkage and association studies provide a promising, unbiased approach for revealing additional regions of the genome associated with UL. In this thesis, I describe results from the first genome-wide linkage and association studies performed in white women affected with UL. A genome-wide linkage study of affected white sister pairs revealed two significant linkage peaks in 10p11 (LOD=4.15) and 3p21 (LOD=3.73) with five suggestive peaks (LOD>2.00) in 2q37, 5p13, 11p15, 12q14, and 17q25. A meta-analysis of genome-wide association results in two independent cohorts of white women

revealed a single nucleotide polymorphism (SNP) with genome-wide significance that is associated with UL diagnosis (rs4247357, P=3.05E-08, odds ratio (OR) =1.299). The candidate SNP is located under the UL linkage peak at 17q25 and is in a large block of linkage disequilibrium (LD) which spans three genes: fatty acid synthase (*FASN*), coiled-coil domain containing 57 (*CCDC57*) and solute carrier family 16, member 3 (*SLC16A3*). *FAS* transcripts and/or protein levels are up-regulated in various neoplasms and have been implicated in tumor cell survival. By tissue microarray immunohistochemistry, we found FAS protein expression elevated three-fold in UL when compared to matched myometrial tissue implicating *FASN* as the first UL risk allele identified in white women by a genome-wide, unbiased approach.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Cynthia Morton, for her support and encouragement over the last five years. Cynthia gave me the opportunity to take the lead on a very important and exciting project in the lab involving cutting edge human genetics. She has encouraged me to become part of the larger human genetics community by attending weekly meetings at the Broad and BWH and presenting my work at multiple local and national meetings. It is obvious that Cynthia truly cares about my development as a scientist and reaching my career goals. I could not have asked for a more inspiring role model and dedicated mentor.

I would also like to thank my dissertation advisory committee, Dr. Raju Kucherlapati, Dr. Kornelia Polyak, and Dr. Brad Quade for their incredibly helpful advice and guidance throughout my graduate school career.

The wonderful members of the Morton lab have played a huge role in my graduate school life and I owe them many thanks for their help along the way. We can debate current topics in science, discuss career goals, troubleshoot experiments and gossip about clothes all in one day. I owe a great deal to Nancy for her advice on several protocols and just about everything else lab-related. And of course, my experience in the Morton lab would not have been the same without my good friend, Jen, who first introduced me to the lab and the field of cytogenetics.

Many of the experiments I present in this thesis would not have been possible without the support of our collaborators. Dr. Rita Cantor and Dr. Ake Lu at UCLA were instrumental in the linkage study analysis and for general guidance on the project. Dr. Dan Chasman and his team at BWH provided an amazing cohort and also a wealth of information on GWAS. Lastly, Dr. Nick Martin and his team also provided an essential cohort and a great deal of valuable experience.

I have been extremely fortunate to find an incredible group of friends in the BBS program. We have shared the highs and lows of graduate school and life over the past five years and have been an endless source of support and laughter for each other.

Most of all, I would like to thank my family for always believing in me. I can honestly say I would not have made it this far without their love and support and they mean the world to me.

For my family

CHAPTER 1:

INTRODUCTION

Uterine leiomyomata (UL), commonly known as fibroids, are benign tumors of the uterine myometrium. Myometrium is made up of smooth muscle tissue and is the thickest layer of tissue in the uterus, flanked by the endometrium and the serosa. UL are classified by tumor location within the myometrial layer: subserosal, intramural, and submucosal. They are hormonally dependent tumors and are not observed prior to puberty. Estrogen, progesterone and other small growth factors (e.g., fibroblast growth factor and transforming growth factor-beta) promote tumor growth and they may regress at menopause. UL are the most prevalent pelvic tumors in women, found in more than 70% of women of reproductive age and approximately 20-25% of women with UL exhibit symptoms (Buttram and Reiter 1981; Cramer and Patel 1990). The most common symptoms of UL include abnormal uterine bleeding and pelvic pressure. Many women with UL present with varying degrees of menorrhagia, or heavy menstrual bleeding. If the tumors exert pressure on nearby organs, such as the bowel or bladder, women may also have problems with bowel movements and urinary frequency. UL cause very serious reproductive issues and complications during pregnancy including recurrent miscarriage, infertility, premature labor, pain, bleeding, fetal malpresentation, and an increased risk of a cesarean section. UL cause severe morbidity but not mortality, likely resulting in limited research funding and treatment options.

The only essentially curative treatment for UL at this time is hysterectomy. Hysterectomies involve surgical removal of the entire uterus and leave the patient unable to bear children. UL is the leading cause for hysterectomy in the United States, accounting for >30% of all hysterectomies and >40% of hysterectomies among women aged 45-64 years (Merrill 2008). Annual health care costs of UL in the United States are

estimated at over 2 billion dollars, most of that cost associated with hysterectomies (Flynn, Jamison et al. 2006). Most women suffering from these tumors are still in their childbearing years so this option is far from ideal. A surgical treatment option that leaves the uterus intact, myomectomy, is also performed frequently. A myomectomy is a surgical procedure in which individual UL are excised and removed from the uterus. Myomectomies are the only treatment for UL that is recommended for women who want to become pregnant; however, approximately half of all myomectomy patients will experience tumor recurrence and 10% will require additional surgery (Fedele, Parazzini et al. 1995).

Less invasive therapies have been developed with limited success. Gonadotropin-releasing hormone (GnRH) analogue therapy is used to induce a low-estrogen environment in the body which decreases UL size. The reduction in size is contingent on continuous exposure to the agonists, and tumors regrow rapidly when therapy has ceased. Long-term use of GnRHa has adverse side effects, specifically for bone density so GnRHa therapy is only used to decrease tumor size temporarily (Matta, Shaw et al. 1989). It is typically administered several months prior to surgery to improve anemia thought to be secondary to menorrhagia, and to allow for a more minimally invasive procedure. Another option, MRI-guided focused ultrasound therapy (FUS), uses high frequency, high energy sound waves to denature proteins in UL cells, thereby causing cell death and tumor destruction. An MRI scanner is used to locate and target UL in the uterus. FUS is a low risk and noninvasive procedure, however, it is not widely available and its effect on future fertility is unknown (Hindley, Gedroyc et al. 2004). Lastly, uterine artery embolization (UAE) is another minimally invasive option that involves

guiding a catheter from a small incision in the groin, through a leg artery, to the arteries in the uterus. Embolic agents are delivered through the catheter to block blood supply to the tumors, which results in UL volume reduction (Hurst, Stackhouse et al. 2000). These therapies do not eliminate tumors but instead relieve symptoms by reducing tumor size. They are much less invasive than hysterectomy, but are not curative and the impact on future fertility is unclear (Nowak 2001). It is evident that more research into the underlying genetic causes of UL will be useful for developing more targeted and less invasive treatments.

Rarely, estimated at 0.1%, UL may progress to their malignant counterpart, uterine leiomyosarcoma (LMS). Expression profiling studies have been pursued in an attempt to understand the relationship between LMS and UL (Figure 1-1). RNA from normal myometrium, UL and LMS was profiled using oligonucleotide microarrays in an attempt to identify genes whose expression levels distinguished the three groups of tissue. The clustering analysis found that normal myometrium and UL were more closely related to each other than to the malignant LMS (Quade, Wang et al. 2004). However, in a follow-up study, a subgroup of UL with unusually cellular pathology and deletions in 1p segregated with the LMS samples, which indicates the profiles of these tumors are more closely related to the malignant tumors than to their benign counterparts and may represent the rare UL that develop into LMS (Christacos, Quade et al. 2006). These profiling experiments have provided a tumor signature for UL and LMS while also highlighting the genetic heterogeneity in these tumors in that different subgroups of UL may exhibit distinct expression profiles that could identify important pathways in UL tumorigenesis and pathobiology.

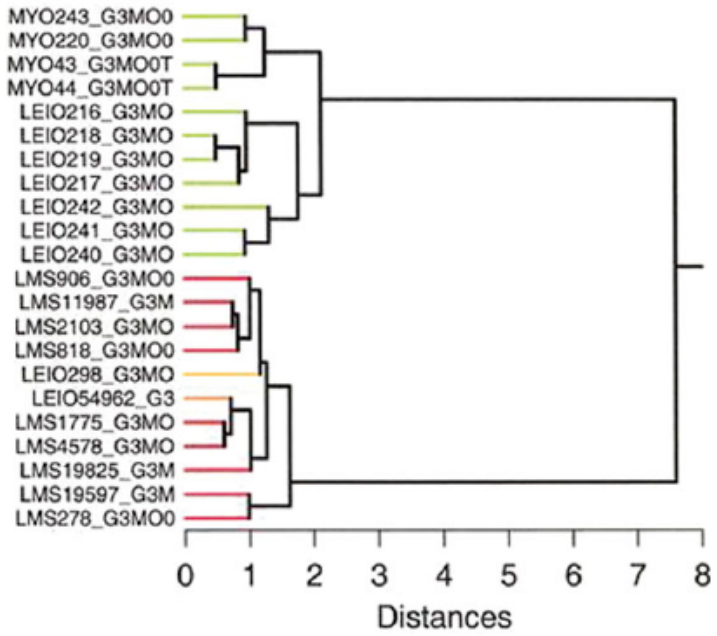


Figure 1-1. Hierarchial clustering illustrates myometrium and UL samples (in green) clustering separately from LMS (in red) and 1p- UL samples (in yellow). Taken from (Christacos, Quade et al. 2006)

Genetic Risk Factors for UL

Several factors predispose women to develop UL. Age, obesity, parity, and race have all been associated with prevalence of UL. Black women are disproportionately affected by UL (Huyck, Panhuysen et al. 2008), with prevalence rates three to four times greater than white women even after controlling for other known risk factors (Marshall, Spiegelman et al. 1997). Affected black women have more and larger UL in comparison to affected white women (Kjerulff, Langenberg et al. 1996). The average age of diagnosis of black women is younger than white women, and black women are more likely to report severe UL symptoms (Huyck, Panhuysen et al. 2008). Genetic events underlying this disparity have yet to be discovered and will undoubtedly provide major insights into UL biology. Further, analyses of twin studies and familial aggregation indicate a genetic component to UL predisposition. First degree relatives of affected probands have a 2.5-fold higher risk of developing UL and monozygotic twins' concordance for UL diagnosis is almost twice that of dizygotic twins' (Vikhlyaeva, Khodzhaeva et al. 1995; Treloar, Do et al. 1998). Similarly, a study of a Finnish cohort found that monozygotic twins' concordance for being hospitalized for UL was twice that of dizygotic twins' (Luoto, Kaprio et al. 2000). These findings support a genetic predisposition to develop UL.

The genetic lesion responsible for a Mendelian disorder involving UL, hereditary leiomyomatosis and renal cell cancer (HLRCC), has been identified through a genome-wide linkage study. HLRCC is caused by mutations in *FH* encoding fumarate hydratase, an enzyme involved in the citric acid cycle. There are approximately 150 families with HLRCC and mutations in *FH*, which are inherited in an autosomal dominant manner

(Toro, Nickerson et al. 2003; Stewart, Glenn et al. 2008). Interestingly, rats with mutations in the tuberous sclerosis 2 gene (*TSC2)* and German shepherd dogs with mutations in the folliculin gene (*BHD)* have very similar phenotypes of both UL and renal cell cancer (Everitt, Wolf et al. 1995; Lingaas, Comstock et al. 2003). Variants in *FH*, *TSC2* and *BHD* do not seem to play a major role in nonsyndromic UL in humans (Gross, Panhuysen et al. 2004).

Genome-wide Genetic Studies

Genome-wide linkage and association studies have led to identification of many genes involved in both rare Mendelian and common diseases. Linkage studies take advantage of related individuals and attempt to identify regions of the genome that are shared between affected family members. Many families with the same disorder can be analyzed and any areas of the genome that are more commonly shared in affected individuals than would be expected by chance are likely to contain the disease-causing mutation. Unfortunately, linkage studies, especially for diseases with significant genetic heterogeneity, often reveal candidate genomic regions that contain over a hundred genes and additional genotyping or sequencing is needed to locate the causal mutation. Genome-wide association studies (GWAS), however, usually identify candidate regions that contain only one or a few genes based on the size of the linkage disequilibrium block, making follow-up experiments potentially more straightforward. In contrast to linkage studies, cohorts used in a GWAS are usually unrelated and much larger in order to provide enough power to detect associations.

In a GWAS, single-nucleotide polymorphism (SNP) allele frequencies across the genome are analyzed and compared in cases and controls. Variants that contribute to the disorder, or are linked to a variant that contributes to the disorder, are expected to have a higher frequency in cases. Conversely, SNPs that do not play a role in affection status are expected to have equal frequencies in cases and controls. In practice, it is unlikely to measure equal allele frequencies and by chance, a single GWAS will identify many SNPs with differential allele frequencies between cases and controls that are not truly associated with the disorder. Because of this, a stringent statistical threshold is used to determine a true association from false positive results. The threshold for a GWAS has been calculated as a p-value less than $5*10^{-8}$, which would be expected by chance only once in 20 GWASs (Altshuler, Daly et al. 2008). Any SNPs that meet or exceed this threshold are considered significantly associated with the disorder, however, many true associations fall below this threshold and are overlooked. SNPs with small effect sizes or very small minor allele frequencies require larger cohorts to reach the power needed to detect a significant association. For most GWASs to date, study population size has been a significant limiting factor for power to detect associations.

Obtaining a substantial population depends largely on the prevalence of the disorder being tested and the ease of stratifying true cases from true controls. UL is very common and most women with UL can be successfully identified using questionnaires. However, women who do not have UL are needed for controls and they are much harder to identify because many women with UL are asymptomatic. The only cohorts large enough to detect associations with UL were stratified based on questionnaires and power was lost due to misclassification of controls because many of the controls are probably

women with asymptomatic UL. To date, only one genome-wide study has been performed for nonsyndromic UL. A genome-wide association study in a Japanese cohort detected three loci significantly associated with UL diagnosis: 10q24.33, 22q13.1, and 11p15.5 (Table 1-1) (Cha, Takahashi et al.). Each locus encompasses several candidate genes, none of which have been implicated in UL biology previously and additional research is required to identify the pathogenetic sequences.

The primary aim of the thesis work presented here is to identify gene variants that predispose white women to develop UL. We performed a genome-wide linkage study and two GWASs in three independent populations of white women. A meta-analysis of the GWAS results revealed one SNP with a genome-wide significant association with UL. This SNP is in a large block of linkage disequilibrium (LD) and is located under one of the UL linkage peaks in 17q25.3. The candidate LD block contains three genes, fatty acid synthase (*FASN)*, coiled-coil domain containing 57 (*CCDC57)* and solute carrier family 16, member 3 (*SLC16A3*). *CCDC57* and *SLC16A3* are presently poorly characterized genes that have no prior association with UL or other tumors. *FASN*, however, has been extensively studied and FAS protein expression is up-regulated in many tumor types (Menendez and Lupu 2007). By immunohistochemistry, we found FAS expression to be elevated three-fold in UL compared to matched, normal myometrium tissue. *FASN* is the first risk allele for UL identified in white women using a genome-wide, unbiased approach. Finding additional pathogenetic sequences that predispose women to UL will provide insight into tumor development and could lead to screening strategies or improved management and therapy.

Table 1-1. GWAS and meta-analysis results of UL diagnosis in a Japanese cohort. Taken from (Cha, Takahashi et al.)

					Case				Control						
Chr.	Chrloc.	SNP	RA	Stage	11	12	22	RAF	11	12	22	RAF	P_{assoc}[a]	P_{het}[b]	OR (95% CI)
10	105,704,389	rs7913069	A	GWAS	22	304	1,281	0.11	5	186	1,237	0.07	7.90×10^{-8}		1.65 (1.37–1.96)
				Rep	31	613	2,795	0.10	23	422	2,799	0.07	8.93×10^{-8}		1.40 (1.23–1.59)
				Combined	53	917	4,076	0.10	28	608	4,036	0.07	8.65×10^{-14}	1.47×10^{-1}	1.47 (1.23–1.75)
22	38,982,819	rs12484776	G	GWAS	291	738	578	0.41	177	647	602	0.35	2.62×10^{-6}		1.29 (1.16–1.43)
				Rep	572	1,623	1,240	0.40	425	1,478	1,341	0.36	2.16×10^{-7}		1.20 (1.12–1.30)
				Combined	863	2,361	1,818	0.41	602	2,125	1,943	0.36	2.79×10^{-12}	2.95×10^{-1}	1.23 (1.11–1.37)
11	193,788	rs2280543	G	GWAS	1,345	252	10	0.92	1093	319	16	0.88	7.16×10^{-7}		1.52 (1.28–1.79)
				Rep	2,829	574	35	0.91	2522	659	63	0.88	5.01×10^{-7}		1.33 (1.19–1.49)
				Combined	4,174	826	45	0.91	3615	978	79	0.88	3.82×10^{-12}	2.06×10^{-1}	1.39 (1.17–1.64)

Recurrent Somatic Mutations in UL

Somatic gene variants involved in UL pathobiology have been discovered by cytogenetic analysis. UL are independent, clonal lesions and approximately 40% have a non-random, tumor specific cytogenetic aberration with several subgroups recognized, including t(12;14)(q14-15;q23-24), del(7)(q22q32), trisomy 12, rearrangements involving 6p21 and 10q22, and deletions of 1p and 3q (Figure 1-2) (Rein, Friedman et al. 1991), (Gross and Morton 2001). Cytogenetic abnormalities have been correlated with tumor size, location, and histology, which indicate that genetic events play a fundamental role in UL biology (Brosens, Deprest et al. 1998),(Christacos, Quade et al. 2006). For example, intramural UL have abnormal karyotypes more frequently than submucosal and subsersol UL (35% vs 12% and 29%) and UL with non-mosaic abnormal karyotypes are significantly larger than UL with normal karyotypes or mosaic abnormal karyotypes (Rein, Friedman et al. 1991). Approximately 20% of UL with abnormal karyotypes have the most common translocation, t(12;14)(q14-15;q23-24). These tumors are usually larger than tumors with a normal karyotype and tumors with del(7)(q22q32), the most common deletion found in UL. Cytogenetic heterogeneity of UL underlies phenotypic differences and supports involvement of different pathways in tumor development.

HMGA2 in UL

Recent cytogenetic-based research of UL has focused on the most common subgroups. Positional cloning of the pathogenetic sequences for two of the subgroups isolated high mobility group proteins (HMG), *HMGA2* at 12q15 in the t(12;14) subgroup

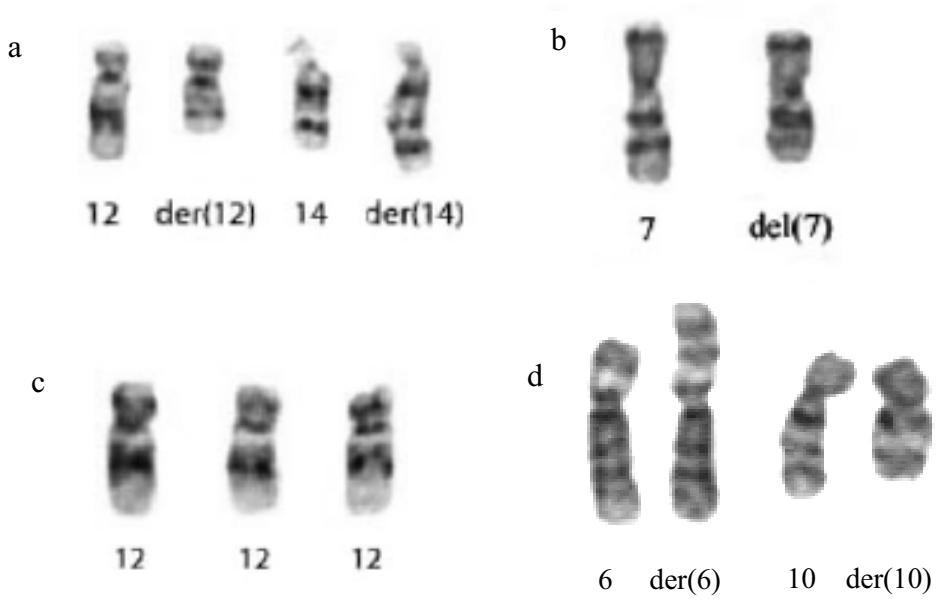
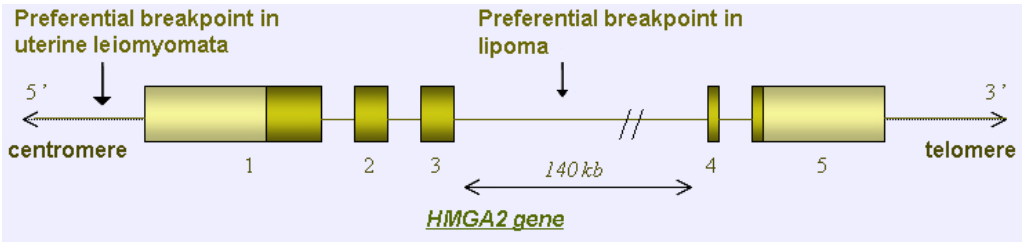


Figure 1-2. Common cytogenetic subgroups found in UL: a) t(12;14)(q14-15;q23-24), b) del(7)(q22q32), c) trisomy 12, d) t(6;10)(q21;q22). Taken from (Gross and Morton 2001).

a



b

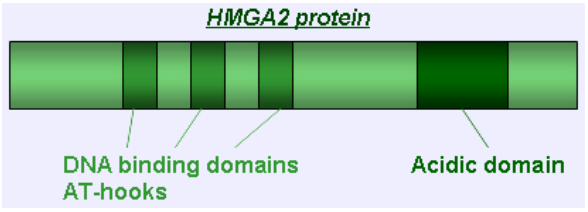


Figure 1-3. Illustration of the *HMGA2* a) gene structure and b) protein structure.

and *HMGA1* at 6p21 (Figure 1-3) (Ashar, Fejzo et al. 1995; Williams, Powell et al. 1997). Increased levels of *HMGA1* and *HMGA2* were found in tumors with 6p21 rearrangements and t(12;14), respectively (Figure 1-4) (Williams, Powell et al. 1997; Gattas, Quade et al. 1999). Trisomy 12 is also a major cytogenetic subgroup in UL and these results taken together suggest that increased expression of HMGA proteins is a key molecular event in UL pathobiology. HMGA1 and HMGA2 are DNA architectural factors which can influence transcription by changing chromatin conformation (Grosschedl, Giese et al. 1994). Both of these proteins have AT hook DNA-binding domains as well as protein binding domains that interact with transcription factors. HMGA2 plays a role in cell differentiation and proliferation in mesenchymal tissue, including myometrium, and abnormal expression of HMGA2 has been linked to several growth-related phenotypes. Transgenic mice that do not express HMGA2 have a *pygmy* phenotype (Zhou, Benson et al. 1995) while mice expressing a truncated form of the protein, which contains the DNA-binding domains but without the regulatory domains, develop an overgrowth phenotype with an unusually high frequency of lipomas (Battista, Fidanza et al. 1999). Recently, a boy with a similar overgrowth phenotype was found to also express a truncated form of *HMGA2*, caused by a chromosomal inversion (Ligon, Moore et al. 2005). Interestingly, several SNPs in the 3' untranslated region (UTR) of *HMGA2* have been identified in GWASs of variation in human height (Lettre, Jackson et al. 2008).

As it is clear that somatic mutations in *HMGA2* are involved in UL development and also that inherited mutations in *HMGA2* can cause overgrowth phenotypes, a candidate gene association study with variants in *HMGA2* was performed to determine if

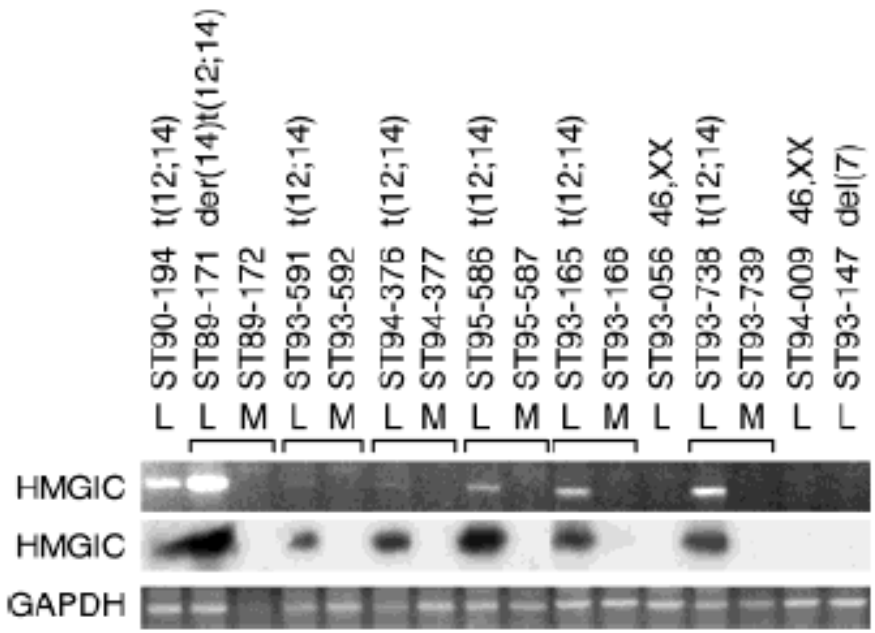
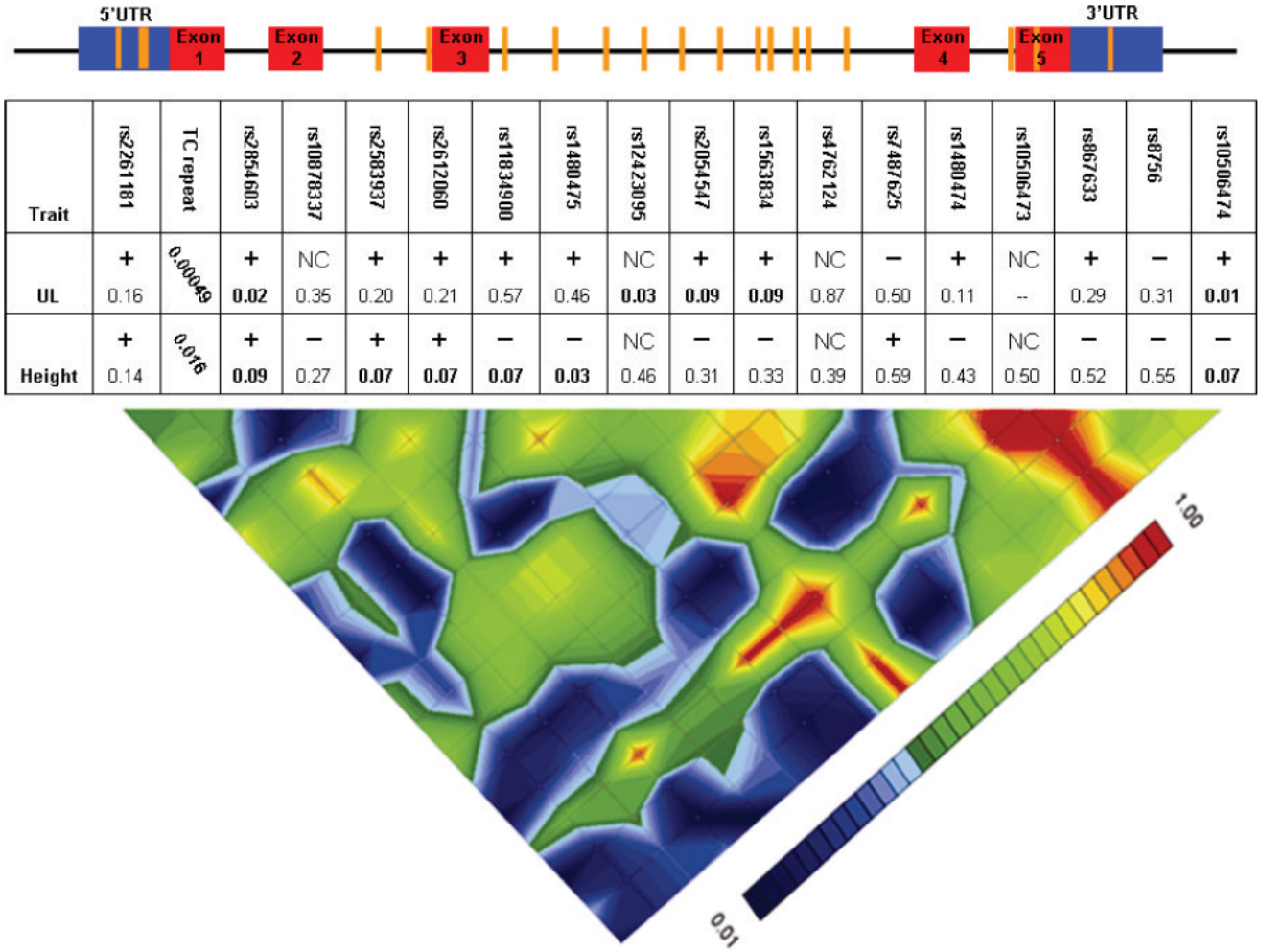


Figure 1-4. RT-PCR amplification of *HMGA2* RNA from six myometrium samples (M) and ten UL (L) with ethidium bromide-stained products (top) and Southern hybridization with a radiolabeled oligonucleotide from *HMGA2* (bottom). Seven UL with t(12;14) clearly demonstrate elevated *HMGA2* expression. Taken from Gattas, Quade et al. 1999.

inherited mutations in *HMGA2* are responsible for predisposition to develop UL. A TC repeat polymorphism in the 5' UTR of *HMGA2* was found to be associated with UL diagnosis in a population of affected white sister pairs (Hodge, K et al. 2009). Although there are approximately 20 alleles of the TC repeat polymorphism in the population, only the allele representing 27 TC repeats was found to be associated with UL diagnosis (Figure 1-5). The second aim of the thesis work presented here is to determine how this variant is affecting *HMGA2* and risk of UL. Because overexpression of *HMGA2* is found in tumors with translocations near the gene and the TC27 repeat is in the 5' UTR, we decided to investigate the possibility that the TC27 repeat affects *HMGA2* expression. One study used luciferase assays to show a significant decrease in *HMGA2* promoter activation when the TC repeat polymorphism is removed (Figure 1-6) (Borrmann, Seebeck et al. 2003). Also, using three TC repeat alleles corresponding to 1, 22, and 36 TC repeats this group demonstrated increased promoter activation with increasing TC repeat number. In an effort to investigate the role of the TC27 repeat, I also performed luciferase assays using the *HMGA2* promoter with varying TC repeat alleles. In 293T cells, there was no significant difference in promoter activation among seven of the most common TC repeat alleles, including TC27. In UL and myometrium cell lines, there was a significant decrease in *HMGA2* promoter activation and large variation in the results suggesting that these cell lines activate the *HMGA2* promoter much differently than 293T cells. An alternative method will need to be developed in order to evaluate the role of the TC27 repeat in *HMGA2* promoter activation in the UL and myometrium cell lines.



Trait	rs2261181	TC repeat	rs2854603	rs10878337	rs2583937	rs2612060	rs11934900	rs1480475	rs12423095	rs2054547	rs1563834	rs4762124	rs7487625	rs1480474	rs10506473	rs867633	rs8756	rs10506474
UL	+ 0.16	0.00049	+ **0.02**	NC 0.35	+ 0.20	+ 0.21	+ 0.57	+ 0.46	NC **0.03**	+ **0.09**	+ **0.09**	NC 0.87	− 0.50	+ 0.11	NC --	+ 0.29	− 0.31	+ **0.01**
Height	+ 0.14	0.016	+ **0.09**	− 0.27	+ **0.07**	+ **0.07**	− **0.07**	− **0.03**	NC 0.46	− 0.31	− 0.33	NC 0.39	+ 0.59	− 0.43	NC 0.50	− 0.52	− 0.55	− **0.07**

Figure 1-5. Candidate gene association results for 17 SNPs and one TC repeat polymorphism across *HMGA2* (marked by orange rectangles) for UL diagnosis and height. LD plot illustrates the D' values between the SNPs. Taken from Hodge, K et al. 2009.

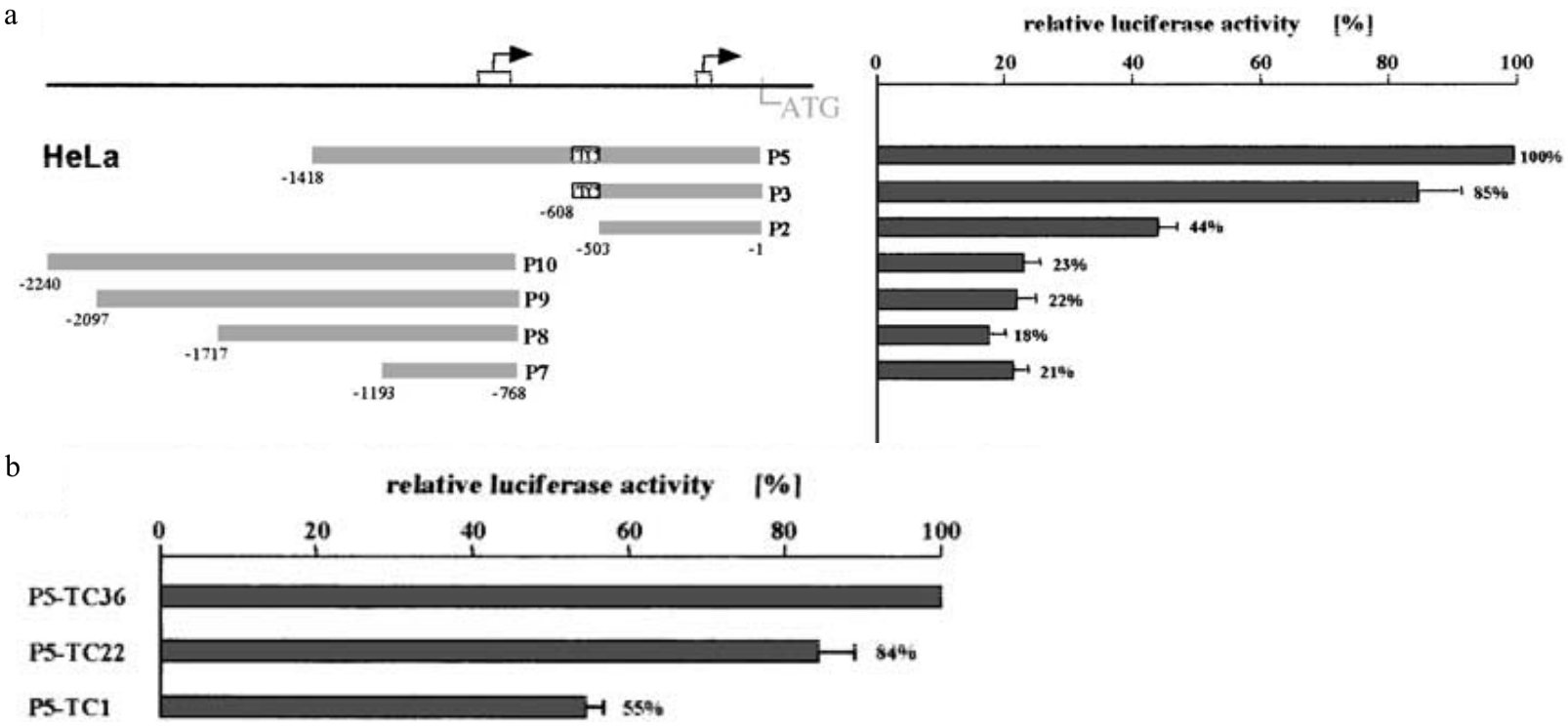


Figure 1-6. Luciferase assay results using several constructs of a) fragments of the *HMGA2* promoter and b) the full *HMGA2* promoter with three TC repeat alleles. Taken from Borrmann, Seebeck et al. 2003.

There is clear evidence for a genetic component to the predisposition to develop UL. Identifying the variants that contribute to UL risk will undoubtedly provide invaluable insights into the biology of these tumors. This thesis describes two such variants, *FASN* and *HMGA2*, which are associated with UL diagnosis and may contribute to UL predisposition in white women. Further study is needed to elucidate the specific role of these variants in UL pathobiology and to evaluate them as possible targets for novel UL therapies.

REFERENCES

Altshuler, D., M. J. Daly, et al. (2008). "Genetic mapping in human disease." Science **322**(5903): 881-8.

Ashar, H. R., M. S. Fejzo, et al. (1995). "Disruption of the architectural factor HMGI-C: DNA-binding AT hook motifs fused in lipomas to distinct transcriptional regulatory domains." Cell **82**(1): 57-65.

Battista, S., V. Fidanza, et al. (1999). "The expression of a truncated HMGI-C gene induces gigantism associated with lipomatosis." Cancer Res **59**(19): 4793-7.

Borrmann, L., B. Seebeck, et al. (2003). "Human HMGA2 promoter is coregulated by a polymorphic dinucleotide (TC)-repeat." Oncogene **22**(5): 756-60.

Brosens, I., J. Deprest, et al. (1998). "Clinical significance of cytogenetic abnormalities in uterine myomas." Fertil Steril **69**(2): 232-5.

Buttram, V. C., Jr. and R. C. Reiter (1981). "Uterine leiomyomata: etiology, symptomatology, and management." Fertil Steril **36**(4): 433-45.

Cha, P. C., A. Takahashi, et al. "A genome-wide association study identifies three loci associated with susceptibility to uterine fibroids." Nat Genet **43**(5): 447-50.

Christacos, N. C., B. J. Quade, et al. (2006). "Uterine leiomyomata with deletions of Ip represent a distinct cytogenetic subgroup associated with unusual histologic features." Genes Chromosomes Cancer **45**(3): 304-12.

Cramer, S. F. and A. Patel (1990). "The frequency of uterine leiomyomas." Am J Clin Pathol **94**(4): 435-8.

Everitt, J. I., D. C. Wolf, et al. (1995). "Rodent model of reproductive tract leiomyomata. Clinical and pathological features." Am J Pathol **146**(6): 1556-67.

Fedele, L., F. Parazzini, et al. (1995). "Recurrence of fibroids after myomectomy: a transvaginal ultrasonographic study." Hum Reprod **10**(7): 1795-6.

Flynn, M., M. Jamison, et al. (2006). "Health care resource use for uterine fibroid tumors in the United States." Am J Obstet Gynecol **195**(4): 955-64.

Gattas, G. J., B. J. Quade, et al. (1999). "HMGIC expression in human adult and fetal tissues and in uterine leiomyomata." Genes Chromosomes Cancer **25**(4): 316-22.

Gross, K. L. and C. C. Morton (2001). "Genetics and the development of fibroids." Clin Obstet Gynecol **44**(2): 335-49.

Gross, K. L., C. I. Panhuysen, et al. (2004). "Involvement of fumarate hydratase in nonsyndromic uterine leiomyomas: genetic linkage analysis and FISH studies." Genes Chromosomes Cancer **41**(3): 183-90.

Grosschedl, R., K. Giese, et al. (1994). "HMG domain proteins: architectural elements in the assembly of nucleoprotein structures." Trends Genet **10**(3): 94-100.

Hindley, J., W. M. Gedroyc, et al. (2004). "MRI guidance of focused ultrasound therapy of uterine fibroids: early results." AJR Am J Roentgenol **183**(6): 1713-9.

Hodge, J. C., T. C. K, et al. (2009). "Uterine leiomyomata and decreased height: a common HMGA2 predisposition allele." Hum Genet **125**(3): 257-63.

Hurst, B. S., D. J. Stackhouse, et al. (2000). "Uterine artery embolization for symptomatic uterine myomas." Fertil Steril **74**(5): 855-69.

Huyck, K. L., C. I. Panhuysen, et al. (2008). "The impact of race as a risk factor for symptom severity and age at diagnosis of uterine leiomyomata among affected sisters." Am J Obstet Gynecol **198**(2): 168 e1-9.

Kjerulff, K. H., P. Langenberg, et al. (1996). "Uterine leiomyomas. Racial differences in severity, symptoms and age at diagnosis." J Reprod Med **41**(7): 483-90.

Lettre, G., A. U. Jackson, et al. (2008). "Identification of ten loci associated with height highlights new biological pathways in human growth." Nat Genet **40**(5): 584-91.

Ligon, A. H., S. D. Moore, et al. (2005). "Constitutional rearrangement of the architectural factor HMGA2: a novel human phenotype including overgrowth and lipomas." Am J Hum Genet **76**(2): 340-8.

Lingaas, F., K. E. Comstock, et al. (2003). "A mutation in the canine BHD gene is associated with hereditary multifocal renal cystadenocarcinoma and nodular dermatofibrosis in the German Shepherd dog." Hum Mol Genet **12**(23): 3043-53.

Luoto, R., J. Kaprio, et al. (2000). "Heritability and risk factors of uterine fibroids--the Finnish Twin Cohort study." Maturitas **37**(1): 15-26.

Marshall, L. M., D. Spiegelman, et al. (1997). "Variation in the incidence of uterine leiomyoma among premenopausal women by age and race." Obstet Gynecol **90**(6): 967-73.

Matta, W. H., R. W. Shaw, et al. (1989). "Long-term follow-up of patients with uterine fibroids after treatment with the LHRH agonist buserelin." Br J Obstet Gynaecol **96**(2): 200-6.

Menendez, J. A. and R. Lupu (2007). "Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis." Nat Rev Cancer **7**(10): 763-77.

Merrill, R. M. (2008). "Hysterectomy surveillance in the United States, 1997 through 2005." Med Sci Monit **14**(1): CR24-31.

Nowak, R. A. (2001). "Identification of new therapies for leiomyomas: what in vitro studies can tell us." Clin Obstet Gynecol **44**(2): 327-34.

Quade, B. J., T. Y. Wang, et al. (2004). "Molecular pathogenesis of uterine smooth muscle tumors from transcriptional profiling." Genes Chromosomes Cancer **40**(2): 97-108.

Rein, M. S., A. J. Friedman, et al. (1991). "Cytogenetic abnormalities in uterine leiomyomata." Obstet Gynecol **77**(6): 923-6.

Stewart, L., G. M. Glenn, et al. (2008). "Association of germline mutations in the fumarate hydratase gene and uterine fibroids in women with hereditary leiomyomatosis and renal cell cancer." Arch Dermatol **144**(12): 1584-92.

Toro, J. R., M. L. Nickerson, et al. (2003). "Mutations in the fumarate hydratase gene cause hereditary leiomyomatosis and renal cell cancer in families in North America." Am J Hum Genet **73**(1): 95-106.

Treloar, S. A., K. A. Do, et al. (1998). "Genetic influences on the age at menopause." Lancet **352**(9134): 1084-5.

Vikhlyaeva, E. M., Z. S. Khodzhaeva, et al. (1995). "Familial predisposition to uterine leiomyomas." Int J Gynaecol Obstet **51**(2): 127-31.

Williams, A. J., W. L. Powell, et al. (1997). "HMGI(Y) expression in human uterine leiomyomata. Involvement of another high-mobility group architectural factor in a benign neoplasm." Am J Pathol **150**(3): 911-8.

Zhou, X., K. F. Benson, et al. (1995). "Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C." Nature **376**(6543): 771-4.

CHAPTER 2

Genome-wide linkage and association analyses in uterine leiomyomata reveal *FASN* as a risk gene

S.L. Eggert[1], K.L. Huyck[4], P. Somasundaram[2], R. Kavalla[2], E.A. Stewart[5], A.T. Lu[6], J.N. Painter[7], G.W. Montgomery[7], S.E. Medland[7], S.A. Treloar[7,8], K.T. Zondervan[9], D.R. Nyholt[10], A.C. Heath[10], P.A.F. Madden[10], L. Rose[11], P.M. Ridker[11,12], D.I. Chasman[11,12], N.G. Martin[7], R.M. Cantor[6], C.C. Morton[2,3,12]

1) Department of Genetics, Harvard Medical School, Boston, MA
2) Departments of Obstetrics, Gynecology and Reproductive Biology and
3) of Pathology, Brigham and Women's Hospital, Boston, MA
4) Department of Medicine, Dartmouth-Hitchcock Medical Center, Lebanon, NH
5) Department of Obstetrics and Gynecology and Surgery, Mayo Clinic and Mayo Medical School, Rochester, MN
6) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA
7) Queensland Institute of Medical Research, Brisbane, Australia
8) The University of Queensland, Centre for Military and Veterans' Health, Herston, Australia
9) Wellcome Trust Centre for Human Genetics and Nuffield Department of Obstetrics and Gynaecology, University of Oxford, UK
10) Washington University School of Medicine, St. Louis, MO
11) Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA
12) Harvard Medical School, Boston, MA

SLE performed all experiments and analyses excluding the linkage analysis. KLH, PS, RK, and EAS recruited and managed samples for FGFF. ATL and RMC performed linkage analysis. JNP, SEM and NGM helped with statistical analysis in Australian cohorts. GWM, SAT, KTZ, DRN, ACH, PAFM, and NGM recruited and managed samples for Australian cohorts. DIC helped with statistical analysis in WGHS cohort. LR, PMR, and DIC recruited and managed samples for WGHS cohort. CCM contributed to overall study design and supervised project.

ABSTRACT

Uterine leiomyomata (UL), the most prevalent pelvic tumors in women of reproductive age, pose a major public health problem given their high frequency, associated morbidities and most common indication for hysterectomies. A genetic component to UL predisposition is supported by analyses of ethnic predisposition, twin studies, and familial aggregation. A genome-wide SNP linkage panel was genotyped and analyzed in 261 white UL sister pair families from the Finding Genes for Fibroids study (FGFF). Two significant linkage regions were detected in 10p11 (LOD=4.15) and 3p21 (LOD=3.73) while five additional linkage regions were identified with LOD scores >2.00 in 2q37, 5p13, 11p15, 12q14 and 17q25. Genome-wide association studies were performed in two independent cohorts of white women and a meta-analysis was conducted. One SNP was identified with a p-value that reached genome-wide significance (rs4247357, P = 3.05E-08, odds ratio (OR) = 1.299). The candidate SNP is under a linkage peak and in a block of linkage disequilibrium in 17q25.3 which spans the genes fatty acid synthase (*FASN)*, coiled-coil domain containing 57 (*CCDC57)* and solute carrier family 16, member 3 (*SLC16A3*). By tissue microarray immunohistochemistry, we found FAS protein expression elevated (3-fold) in UL when compared to matched myometrial tissue. *FAS* transcripts and/or protein levels are up-regulated in various neoplasms and implicated in tumor cell survival. *FASN* represents the first UL risk allele identified in white women by a genome-wide, unbiased approach and opens a path to management and potential therapeutic intervention.

INTRODUCTION

Uterine leiomyomata (UL), commonly known as fibroids, are benign tumors of the uterine myometrium. They represent the most prevalent pelvic tumors in women, found in more than 70% of women of reproductive age (Cramer and Patel 1990). Approximately 20-25% of women with UL exhibit symptoms including menorrhagia, infertility, pelvic pain and a range of complications during pregnancy (Buttram and Reiter 1981). The leading cause for hysterectomy in the United States, UL account for >30% of all hysterectomies and >40% of hysterectomies among women aged 45-64 years (Lepine, Hillis et al. 1997). Annual health care costs of UL are estimated at over two billion dollars, most of that cost associated with hysterectomies (Flynn, Jamison et al. 2006). Although UL pose a major public health problem, little is known about the molecular basis for these tumors and treatment options are limited.

Genes involved in UL have been discovered by cytogenetic analysis. Approximately 40% of UL have a non-random cytogenetic aberration and several subgroups are recognized, including t(12;14)(q14-15;q23-24), del(7)(q22q32), trisomy 12, rearrangements involving 6p21 and 10q22, and deletions of 1p and 3q (Rein, Friedman et al. 1991; Gross and Morton 2001). Cytogenetic abnormalities have been correlated with tumor size, location, and histology, which indicate that genetic events play a fundamental role in UL biology (Rein, Friedman et al. 1991; Brosens, Deprest et al. 1998; Christacos, Quade et al. 2006). Cytogenetic heterogeneity of UL underlies phenotypic differences and supports involvement of different pathways in tumor development.

Several factors predispose women to develop UL. Age, obesity, parity, and race have all been associated with prevalence of UL. Black women are disproportionately affected by UL (Huyck, Panhuysen et al. 2008), with incidence and prevalence rates at least three times greater than that in white women even after controlling for other known risk factors (Marshall, Spiegelman et al. 1997). Further, analyses of twin studies and familial aggregation indicate a genetic component to UL predisposition; first degree relatives of affected women have a 2.5-fold higher risk of developing UL and monozygotic twins' concordance for UL diagnosis is almost twice that of dizygotic twins' (Vikhlyaeva, Khodzhaeva et al. 1995; Treloar, Do et al. 1998). Similarly, a study of a Finnish cohort found that monozygotic twins' concordance for being hospitalized for UL was twice that of dizygotic twins' (Luoto, Kaprio et al. 2000). These findings support a genetic predisposition to develop UL but no genome-wide study of UL in white women has been reported. Several candidate gene association studies have been performed with limited success, although variants in the 5' UTR of *HMGA2*, a gene involved in recurrent cytogenetic aberrations of UL and known to play a primary role (Schoenberg Fejzo, Ashar et al. 1996), have been associated with UL diagnosis in a cohort of white sister pairs (Hodge, K et al. 2009). Finding additional pathogenetic sequences that predispose women to UL will provide insight into tumor development and could lead to screening strategies or improved management and therapy.

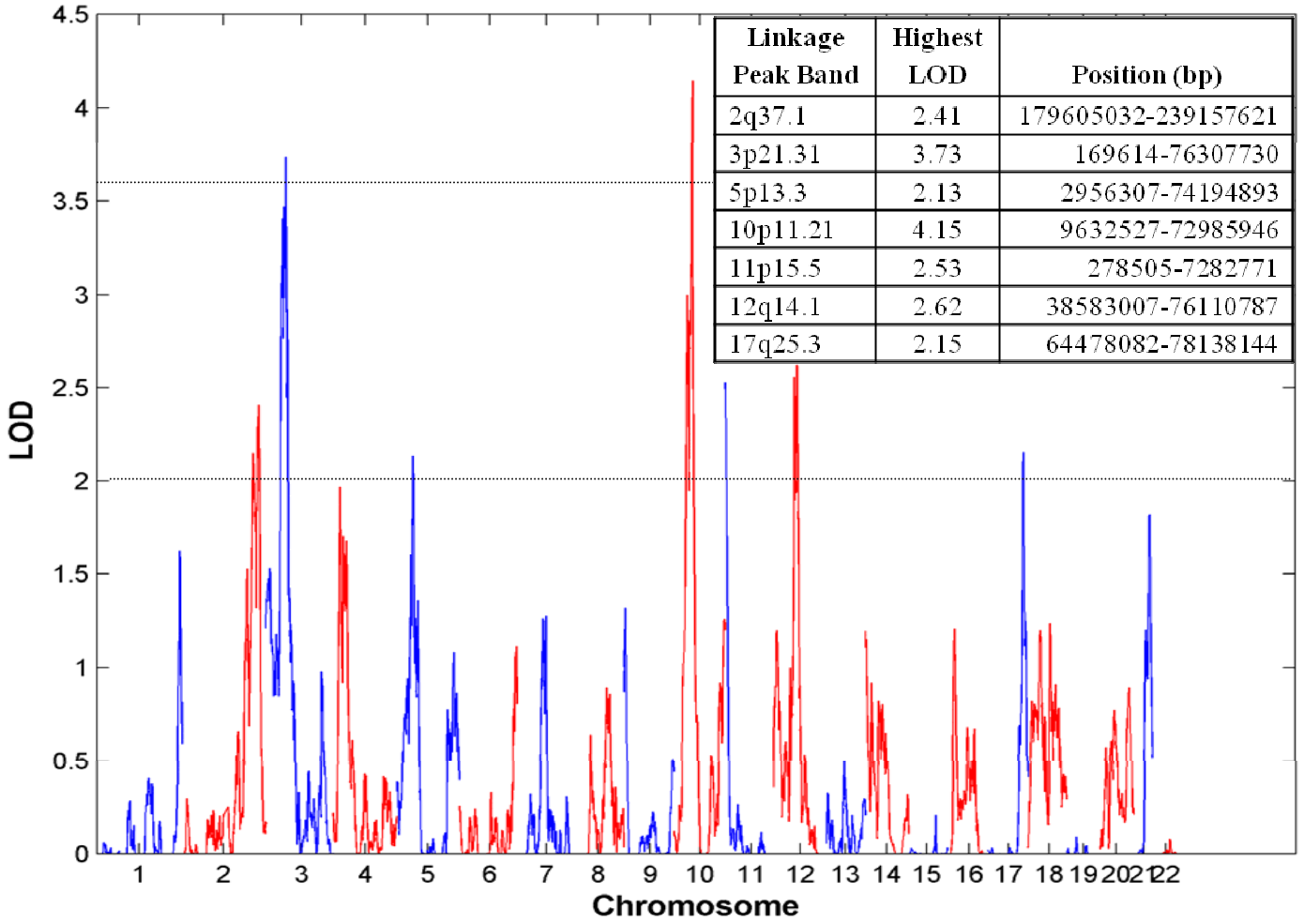
RESULTS

Linkage Study

Linkage analysis with the Finding Genes for Fibroids (FGFF) population (see Methods) revealed two peaks with highly significant genome-wide LOD scores (>3.6) and five peaks with suggestive LOD scores (>2.0) (Figure 2-1). The highest LOD scores found are at 10p11 (LOD=4.15) and at 3p21 (LOD=3.73). Both linkage regions are around 35 Mb and contain hundreds of genes. Of note, *HMGA2* resides within the linkage region at 12q14 with a suggestive LOD score of 2.62 (Figure 2-2).

Association Analyses and Meta-Analysis

Genome-wide association studies (GWAS) were undertaken with two independent cohorts of white women, the Women's Genome Health Study (WGHS) cohort and an Australian cohort (see Methods). Analysis of the WGHS cohort revealed 45 SNPs with p-values less than 10^{-4} (Table 2-1). Although none of the p-values from this analysis are considered significant to identify a genome-wide association, the quantile-quantile plot of the results provides evidence that there are more SNPs with small p-values than expected by chance (Figure 2-3a). In the Australian analysis, 25 SNPs were identified with p-values less than 10^{-4} and the quantile-quantile plot reveals a small increase in low p-values although it is less striking than the WGHS results (Table 2-2, Figure 2-3b). Meta-analysis was performed on the set of 344,655 genotyped SNPs from the WGHS and Australian cohorts using an inverse-variance weighted method in METAL (http://www.sph.umich.edu/csg/abecasis/Metal/index.html). One SNP, rs4247357, reached genome-wide significance and is considered significantly associated



Linkage Peak Band	Highest LOD	Position (bp)
2q37.1	2.41	179605032-239157621
3p21.31	3.73	169614-76307730
5p13.3	2.13	2956307-74194893
10p11.21	4.15	9632527-72985946
11p15.5	2.53	278505-7282771
12q14.1	2.62	38583007-76110787
17q25.3	2.15	64478082-78138144

Figure 2-1. Summary of FGFF sib-pair linkage analysis by chromosome. Two peaks have significant LOD scores (>3.6) and five peaks have suggestive LOD scores (>2.0). Linkage peak boundaries and the highest LOD score found under each peak are defined in the inserted table.

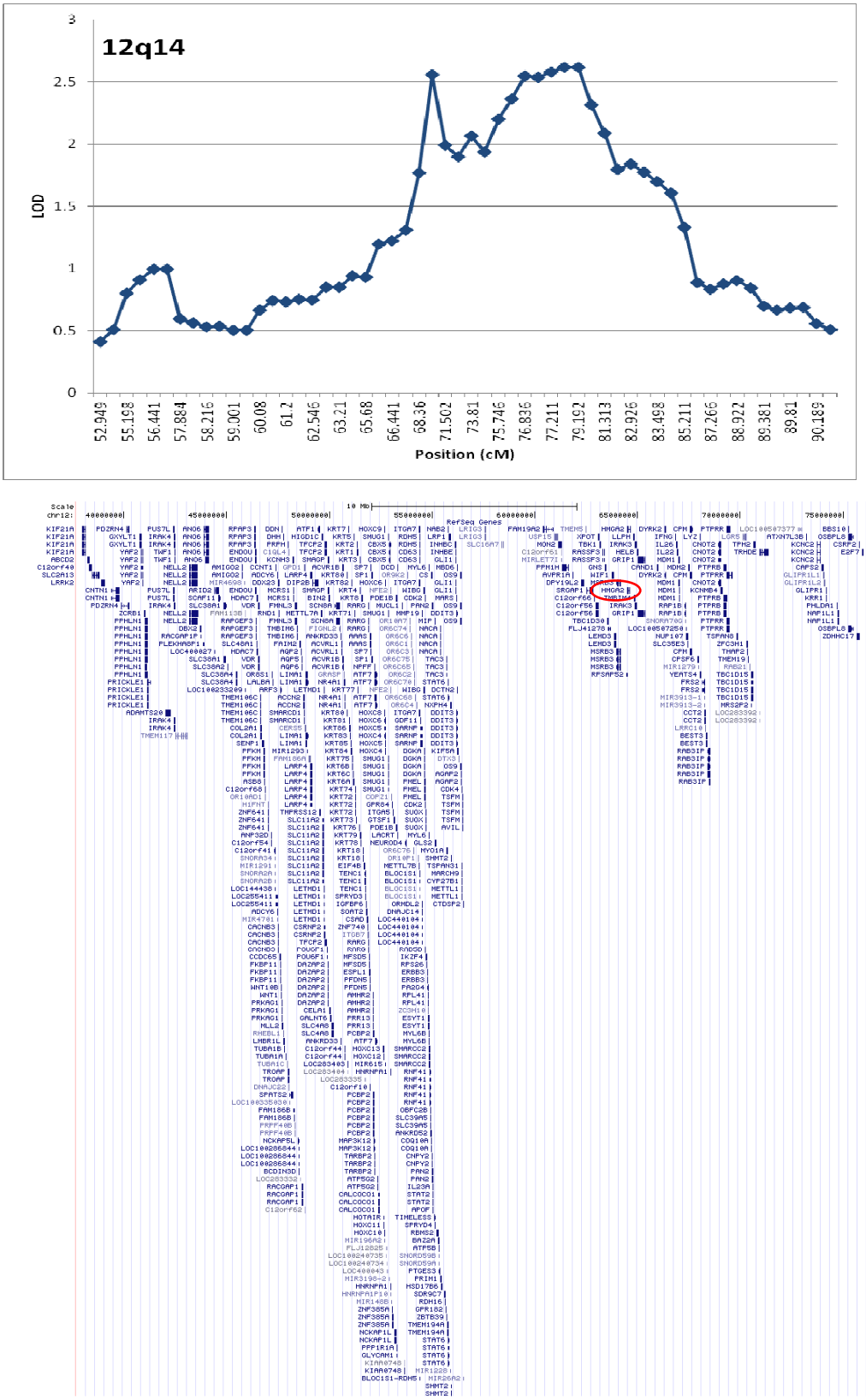


Figure 2-2. FGFF linkage peak at 12q14 with *HMGA2* circled in red.

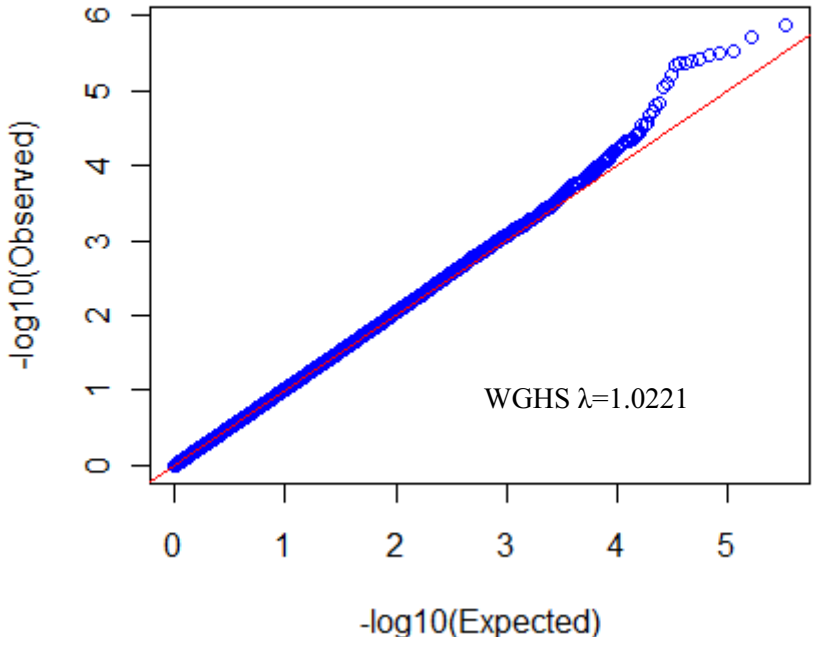
Table 2-1. WGHS genome-wide association analysis results ($P<10^{-4}$)

CHR	SNP	Position (bp)	F_A	F_U	A2	CHISQ	P-value
1	rs2268179	22287371	0.2091	0.1588	C	23.37	1.34E-06
17	rs4247357	77760277	0.467	0.4014	C	22.64	1.95E-06
17	rs6502057	77675348	0.4628	0.3977	G	21.78	3.06E-06
15	rs2903332	78340261	0.3124	0.2542	G	21.75	3.10E-06
17	rs11077969	77678921	0.4655	0.4017	G	21.31	3.91E-06
17	rs7502078	77725167	0.4678	0.4044	C	21.24	4.04E-06
17	rs7213172	77678991	0.4657	0.4023	A	21.16	4.23E-06
17	rs8080423	77695567	0.4672	0.4039	A	21.1	4.36E-06
17	rs7406163	77679677	0.4772	0.4128	A	21.06	4.46E-06
17	rs7221544	77680625	0.4664	0.4042	A	20.41	6.26E-06
9	rs11139665	84451020	0.05295	0.03042	A	19.98	7.84E-06
22	rs732110	26345321	0.1166	0.08168	A	19.73	8.90E-06
5	rs30523	132138941	0.1516	0.1125	G	18.71	1.52E-05
1	rs6693503	19859880	0.4135	0.3555	G	18.61	1.61E-05
14	rs1951054	25147247	0.3004	0.3575	A	18.33	1.86E-05
1	rs2012235	68685342	0.2332	0.2866	A	18.11	2.09E-05
4	rs2333255	176819759	0.4289	0.4874	C	17.55	2.80E-05
1	rs12563321	26539088	0.2554	0.2074	G	17.5	2.87E-05
1	rs9729637	5468038	0.09182	0.06265	C	17.45	2.95E-05
14	rs1257670	98562725	0.3394	0.2867	A	17.06	3.63E-05
1	rs12117956	27370168	0.305	0.3598	A	16.89	3.97E-05
9	rs7855598	137454075	0.2262	0.1813	A	16.84	4.06E-05
1	rs4846689	219444057	0.1156	0.0832	A	16.71	4.36E-05
6	rs6903101	138567539	0.04625	0.02687	A	16.69	4.39E-05
17	rs917538	14975565	0.4538	0.3978	C	16.6	4.61E-05
1	rs6687674	27409687	0.362	0.4182	G	16.58	4.67E-05
14	rs2284230	77085649	0.175	0.2217	G	16.5	4.88E-05
10	rs3824700	26395911	0.4953	0.4388	A	16.48	4.92E-05
11	rs7945105	131162979	0.3653	0.4212	G	16.48	4.93E-05
10	rs1521032	26393597	0.4946	0.4385	A	16.3	5.41E-05
1	rs10902742	26548843	0.2372	0.1923	C	16.13	5.91E-05
10	rs11593128	26392957	0.4946	0.4385	A	16.09	6.04E-05
10	rs3936497	28619974	0.1034	0.0734	A	16	6.34E-05
1	rs1454356	189400411	0.3495	0.4044	A	15.97	6.42E-05
9	rs1556047	112083819	0.5121	0.4565	A	15.86	6.83E-05
2	rs12692335	7524145	0.2483	0.203	G	15.84	6.90E-05

Table 2.1 (Continued)

8	rs10955841	119015721	0.3718	0.3198	A	15.67	7.52E-05
1	rs9786944	5471098	0.08914	0.06186	G	15.5	8.24E-05
6	rs6926282	112968480	0.128	0.09509	G	15.43	8.55E-05
21	rs13049184	21818485	0.4448	0.4998	G	15.4	8.70E-05
12	rs12579612	93189923	0.2426	0.1984	A	15.39	8.75E-05
10	rs1521033	26422356	0.494	0.4395	G	15.36	8.89E-05
6	rs627240	81042603	0.3834	0.4387	C	15.33	9.04E-05
15	rs11636483	96937674	0.2507	0.206	A	15.25	9.44E-05
16	rs16953111	78954425	0.09987	0.07114	G	15.17	9.83E-05

a



b

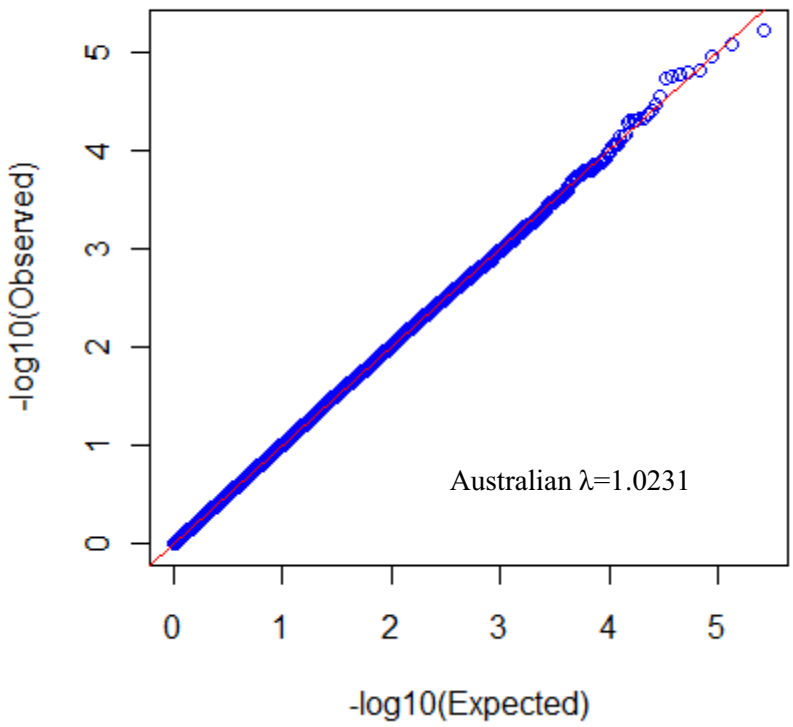


Figure 2-3. Quantile-quantile plot of (a) WGHS GWAS results and (b) Australian GWAS results. P-values observed in the studies are compared to p-values expected under the null hypothesis.

Table 2-2. Australian genome-wide association analysis results ($P<10^{-4}$)

CHR	SNP	Position (bp)	F_A	F_U	A2	CHISQ	P-value
8	rs10504743	82951580	0.04752	0.09918	G	20.43	6.18E-06
1	rs10779614	212656537	0.1777	0.2574	G	19.84	8.43E-06
1	rs6703314	104976389	0.3244	0.2397	A	19.29	1.13E-05
12	rs6539579	80696724	0.2562	0.3418	A	18.7	1.53E-05
3	rs4128782	89320422	0.468	0.5607	G	18.57	1.64E-05
7	rs1829993	77698210	0.469	0.5615	G	18.49	1.71E-05
5	rs10515600	147316068	0.3089	0.3975	C	18.45	1.74E-05
17	rs11079098	49220927	0.3295	0.4189	A	18.28	1.90E-05
2	rs750132	218548607	0.1508	0.09262	G	17.52	2.84E-05
17	rs3785655	14189892	0.2996	0.3844	T	17.15	3.46E-05
6	rs9389508	137842904	0.1405	0.08525	C	16.87	4.00E-05
18	rs4542757	48452722	0.3068	0.391	T	16.72	4.32E-05
7	rs6967325	77678568	0.407	0.323	A	16.56	4.72E-05
10	rs2082988	128254551	0.3409	0.4262	G	16.54	4.77E-05
20	rs6116201	4033758	0.1498	0.218	A	16.47	4.93E-05
20	rs10485664	38064192	0.2293	0.1607	A	16.48	4.93E-05
17	rs1859906	14191413	0.3688	0.4549	G	16.47	4.95E-05
9	rs11144978	78415148	0.3089	0.232	A	16.36	5.25E-05
1	rs1808973	112289357	0.4122	0.4975	T	15.83	6.93E-05
7	rs798332	77746864	0.3946	0.3131	T	15.78	7.11E-05
3	rs7429534	89273999	0.3564	0.4402	C	15.74	7.28E-05
6	rs6570048	136240287	0.4287	0.5131	T	15.41	8.64E-05
1	rs3006009	242697070	0.3533	0.4361	T	15.41	8.65E-05
1	rs11120315	212653320	0.1126	0.1721	C	15.36	8.90E-05
12	rs12228394	19482075	0.01756	0.04836	C	15.27	9.32E-05

with UL status (Table 2-3). Five additional SNPs in the same location on chromosome 17 were identified with p-values less than 10^{-6}. The quantile-quantile plot of the meta-analysis results clearly shows that this group of SNPs have lower p-values than would be expected under the null hypothesis (Figure 2-4). The candidate SNPs on chromosome 17 are located in a large linkage disequilibrium (LD) block which contains three genes, fatty acid synthase (*FASN*), coiled-coil domain containing 57 (*CCDC57*) and solute carrier family 16, member 3 (*SLC16A3*) (Figure 2-5, Figure 2-6). Interestingly, this LD block lies under the FGFF linkage peak at 17q25 (Figure 2-7).

FAS Protein Expression

Little is known about CCDC57 or SLC16A3, however, FAS has been associated in several cancers and is often up-regulated in cancer tissue (Menendez and Lupu 2007). In order to investigate this finding in UL, we stained UL and myometrial tissue with a FAS antibody. FAS immunostaining revealed a three-fold increase in FAS expression in UL compared to matched, normal myometrium (Figure 2-8). An increase in UL FAS expression was seen in 25 out of 33 (~76%) matched samples (Figure 2-9). Stratifying matched samples by rs4247357 genotype shows an increase in FAS expression in myometrium and UL with the major allele compared to the minor allele. Increased expression in myometrium samples with the major allele is not statistically significant but the increase in UL with the major allele is about two-fold and is significant (Figure 2-10a). When all UL samples are combined, the same pattern emerges but the increase in FAS expression in UL with the major allele is about 50% higher than UL with the minor allele (Figure 2-10b). Interestingly, in both matched and combined analyses, FAS

Table 2-3. Top SNP from meta-analysis results of WGHS and Australian cohorts ($P<5\times10^{-8}$)

SNP	Position	Cohorts	MAF	OR (95% CI)	P-value
rs4247357	17q25.3 at 77760277 in *CCDC57*	WGHS	0.467	1.307 (1.170-1.459)	1.95E-06
		Australian	0.432	1.282 (1.079-1.523)	4.49E-03
		Meta-analysis		1.299 (1.184-1.426)	3.05E-08

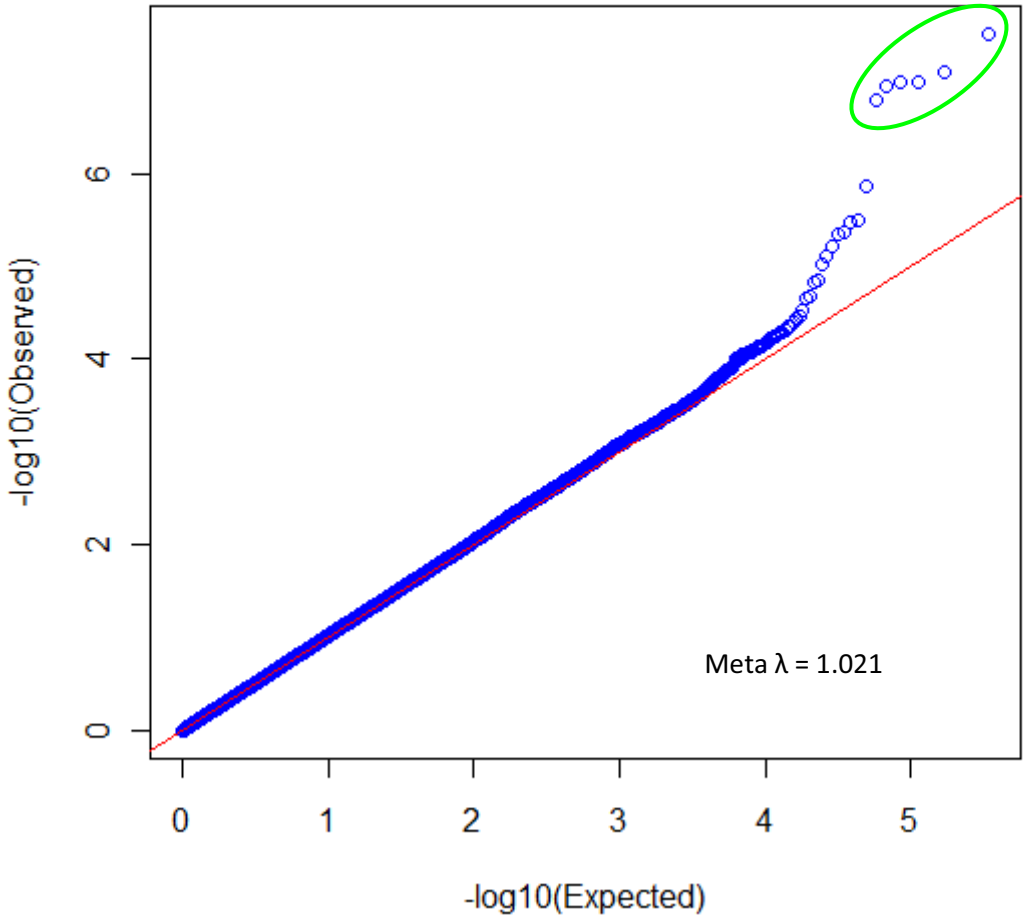


Figure 2-4. Quantile-quantile plot of meta-analysis results. P-values observed in the study are compared to p-values expected under the null hypothesis. Top SNPs in 17q25.3 region are circled in green.

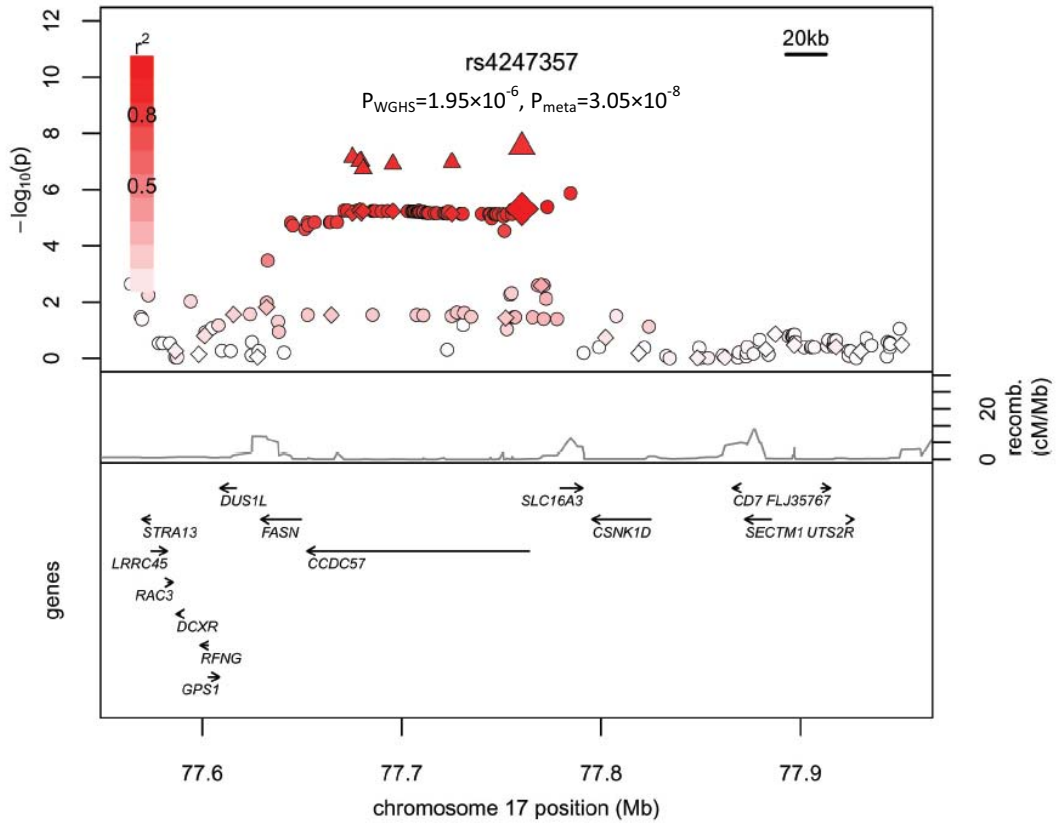


Figure 2-5. Candidate region on chromosome 17 containing significantly associated markers. P-values for WGHS genotyped SNPs are indicated by diamonds, p-values for WGHS imputed SNPs are indicated by circles and p-values for SNPs included in the meta-analysis are indicated by triangles.

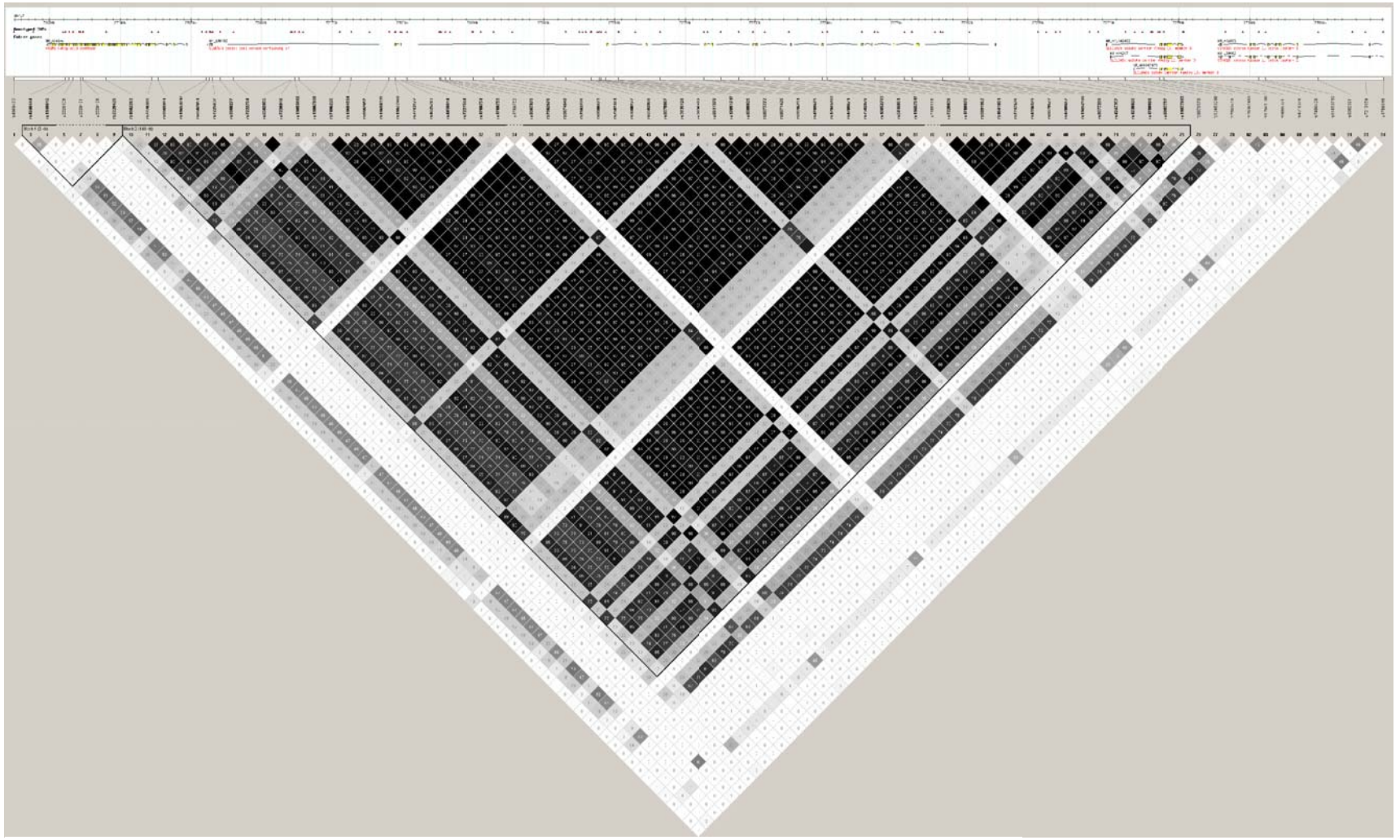


Figure 2-6. HapMap generated LD plot of candidate region on chromosome 17.

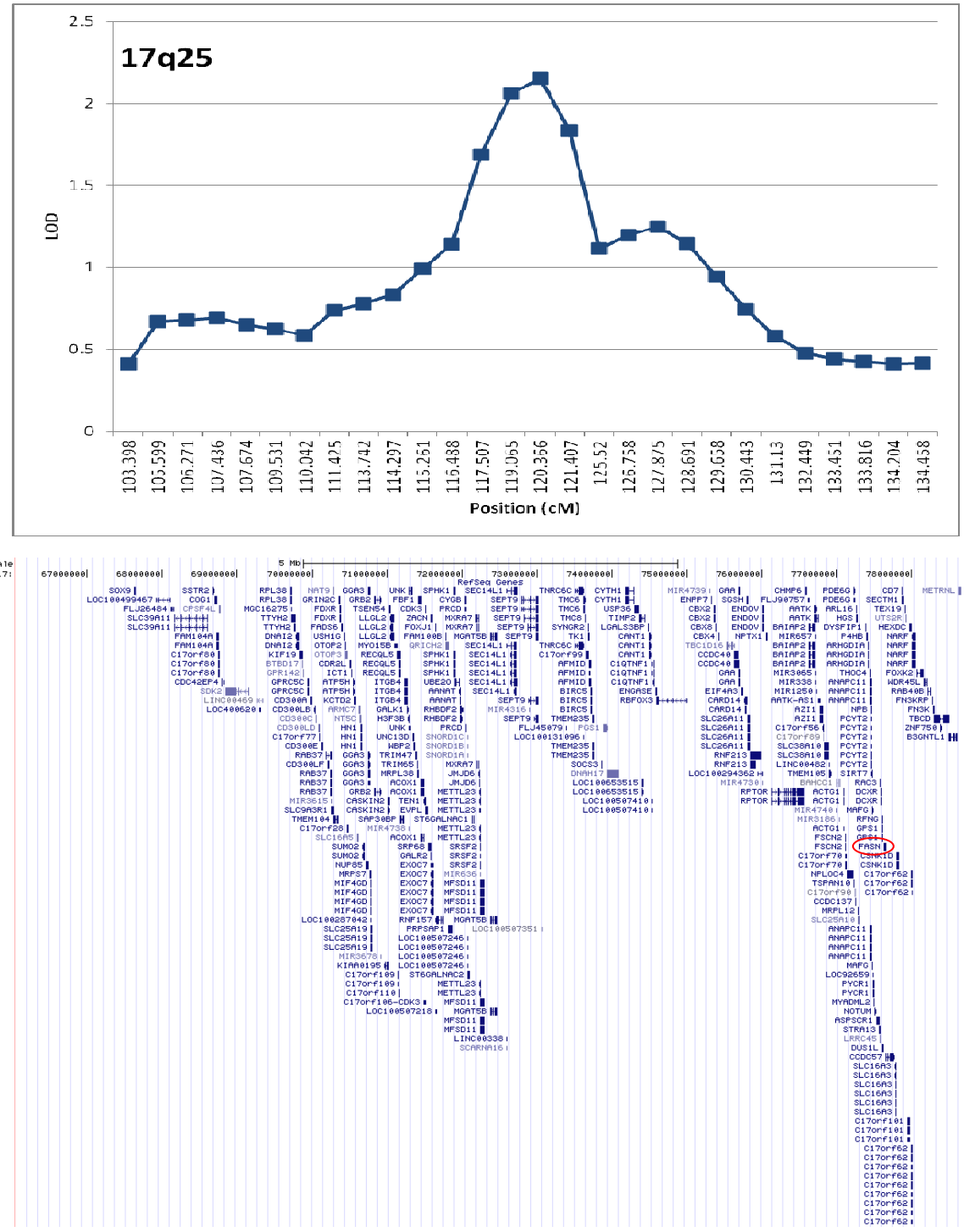


Figure 2-7. FGFF linkage peak at 17q25 with *FASN* circled in red.

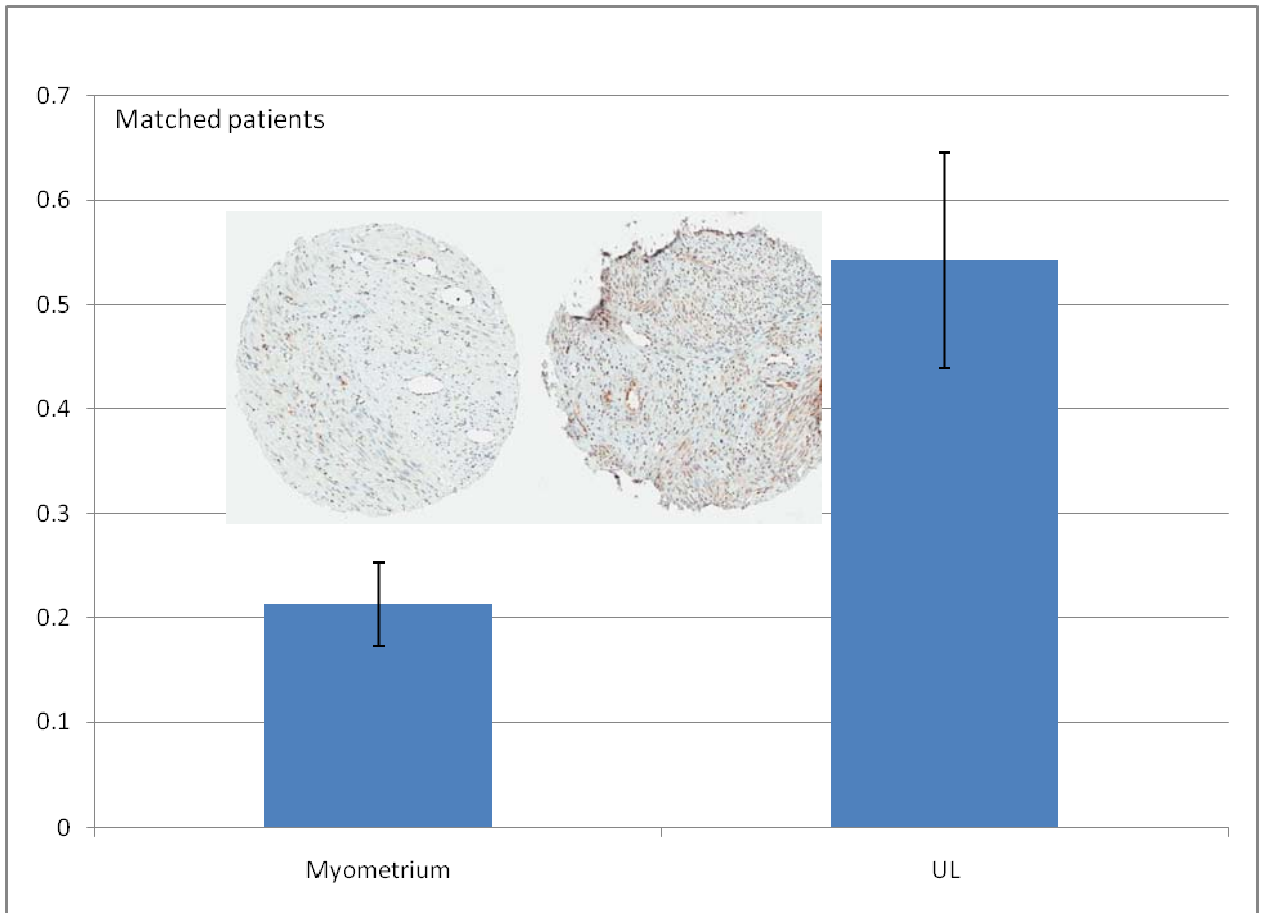


Figure 2-8. FAS protein expression in myometrium and UL from matched samples. Representative tissue cores are shown from myometrium (left) and a UL (right).

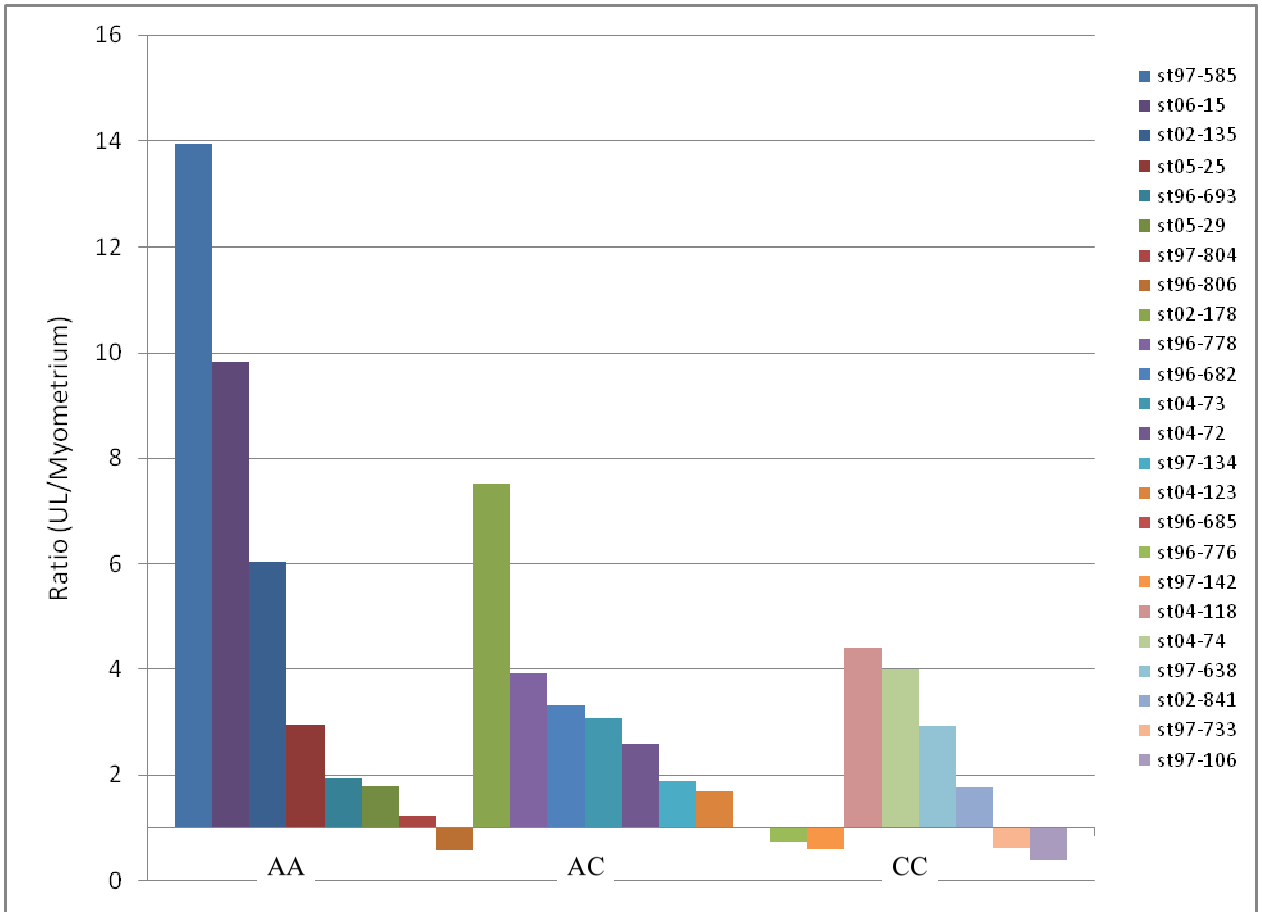
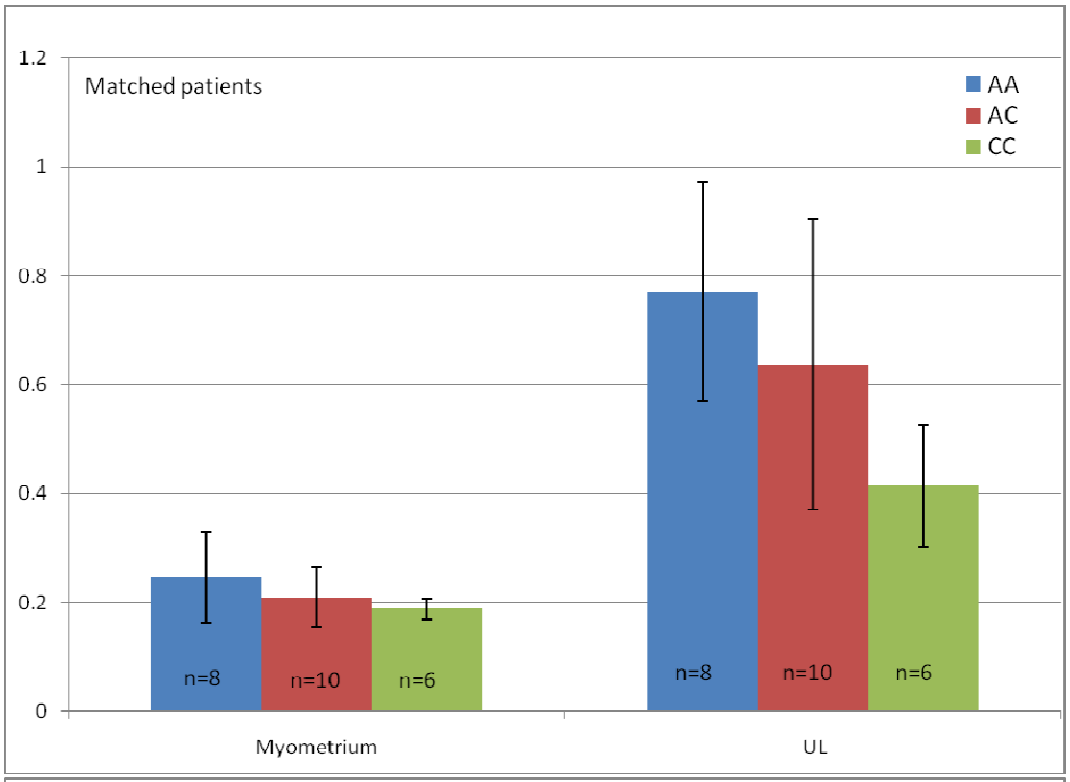


Figure 2-9. Ratio of FAS protein expression in matched UL/myometrium samples grouped by rs4247357 genotype. Each bar represents the ratio from one woman.

a



b

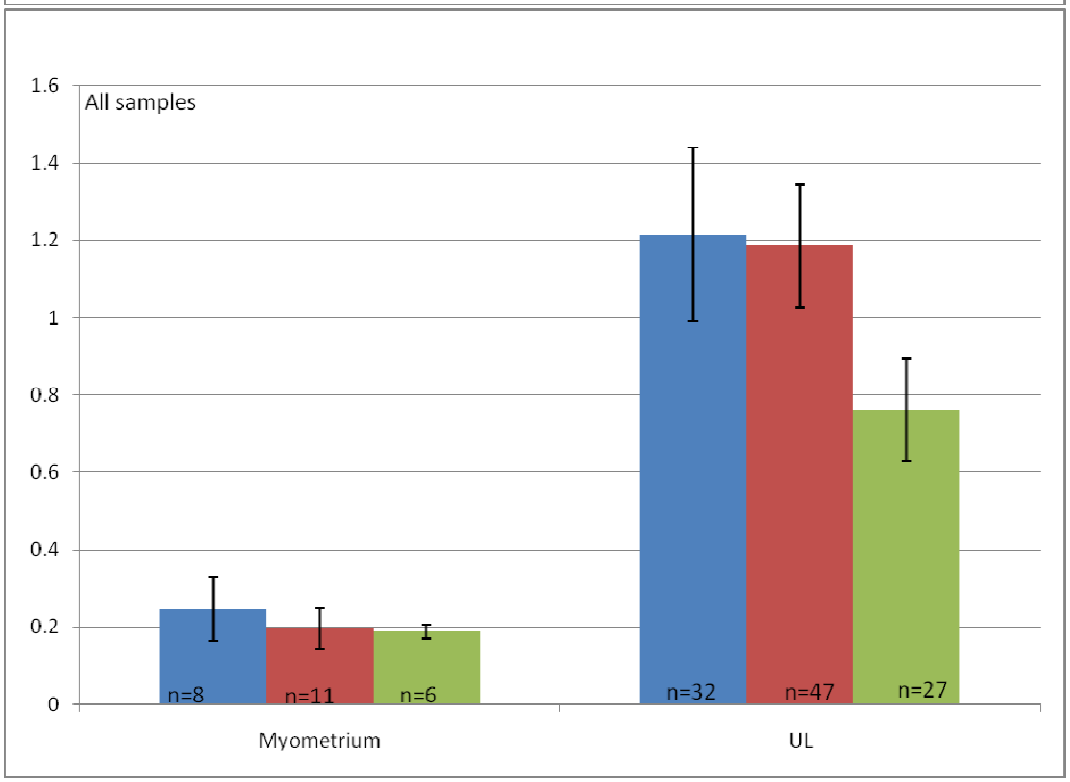


Figure 2-10. FAS protein expression in myometrium and UL from (a) matched samples and (b) all samples stratified by rs4247357 genotype.

expression in both myometrium and UL samples heterozygous for major and minor alleles was at a level between that of samples homozygous for either the major or minor alleles.

mRNA Expression

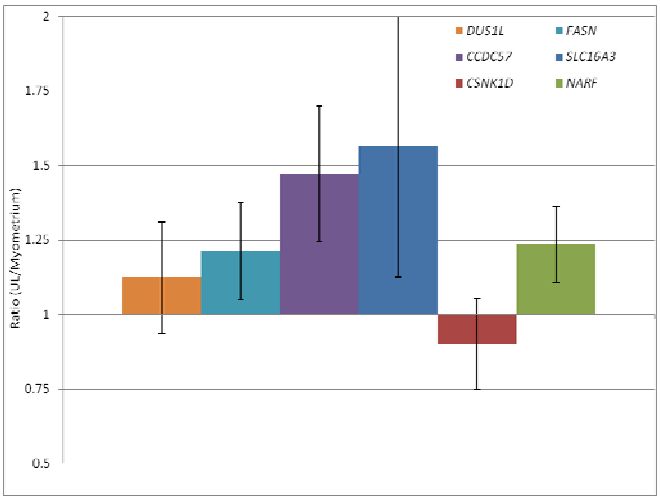
Analysis of mRNA expression levels in myometrium and UL by microarray of genes under the linkage peak on chromosome 17 revealed no genes with significant differential expression between women with the major (n=6) and minor alleles (n=3) of rs4247357. *FASN*, *CCDC57* and *SLC16A3* were found not to be expressed at a detectable level in the myometrium and UL samples analyzed. However, a more comprehensive analysis of 54,000 probes across the genome revealed several genes significantly upregulated and downregulated in UL with the minor allele compared to UL with the major allele (Table 2-4).

Targeted mRNA expression analysis by qPCR detected no substantial expression differences of matched myometrium and UL samples for *FASN*, *CCDC57*, *SLC16A3* and three genes located directly nearby but outside of the candidate LD block, *DUS1L*, *CSNK1D* and *NARF* (Figure 2-11a). Expression of *FASN, CCDC57* and *SLC16A3* is slightly higher in the matched UL samples but expression of *DUS1L* and *NARF* is also higher in the matched UL samples and the variation between samples is relatively large. It is unclear whether this expression difference has biological significance; however, it is clear that mRNA expression of *FASN* between myometrium and matched UL samples is not concordant with FAS protein expression as the same samples were used in both studies. This finding is not surprising as the correlation between mRNA and protein levels can be poor, with some studies finding that only 40% of variation in protein.

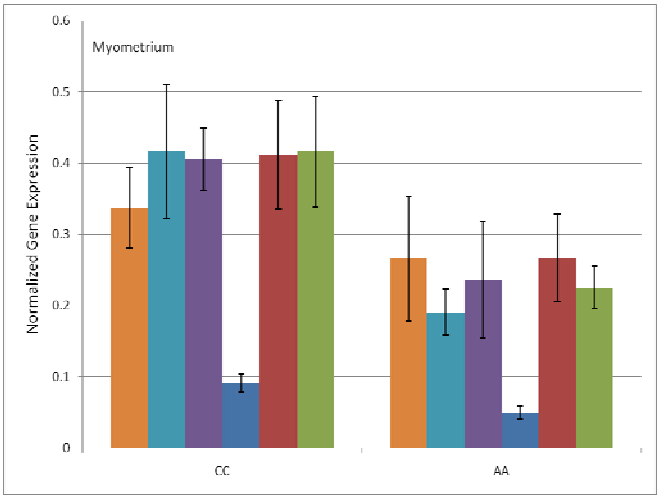
Table 2-4. mRNA of genes found to be upregulated or downregulated in UL with the minor allele of rs4247357 compared to UL with the major allele.

Gene	Location	Fold Change	Direction	P-value
MMS19	10q24.1	1.558	up	7.45E-07
SH3BP4	2q37.2	1.804	down	8.42E-07
TUBB2A	6p25.2	2.658	down	1.86E-06
NLN	5q12.3	1.617	down	2.19E-06
FAM126A/DRCTNNB1A	7p15.3	1.699	down	6.49E-06
STX7	6q23.2	2.014	down	6.88E-06
PDS5B/KIAA0979	13q13.1	1.753	down	7.25E-06
CR613961	1p36.11	2.655	down	7.96E-06
CR749816	1p21.2	1.817	down	8.49E-06
PMEPA1	20q13.31	2.284	down	1.16E-05
MAP3K7IP3	Xp21.2	2.581	down	1.17E-05
HOOK3	8p11.21	1.650	down	1.46E-05
C12orf47	12q24.12	1.728	up	1.47E-05
ANKRD27	19q13.11	1.870	down	1.55E-05
RAB6C	2q21.1	1.407	down	1.71E-05
C6orf162	6q15	1.585	up	1.82E-05
PICALM	11q14.2	1.884	down	1.83E-05
c3orf19	3p25.1	1.310	up	2.07E-05
ZSIG13/FZD4	11q14.2	2.527	down	2.15E-05
PIAS1	15q23	1.685	up	2.36E-05
FADS1	11q12.2	1.948	down	2.38E-05
PHF21A/KIAA1696	11p11.2	1.343	up	2.45E-05
RBM15	1p13.3	1.810	down	2.65E-05
DIS3	13q22.1	1.565	down	2.78E-05

a



b



c

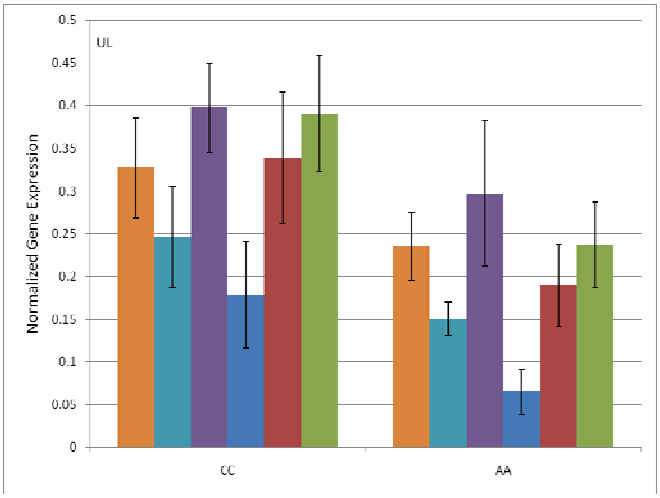


Figure 2-11. mRNA expression by qPCR of *FASN*, *CCDC57, SLC16A3* and three genes in direct proximity (*DUS1L, CSNK1D, NARF*) in (a) myometrium and matched UL samples (b) myometrium with the major (AA) and minor (CC) allele of rs4247357 and (c) UL with the major (AA) and minor (CC) allele of rs4247357.

expression can be explained by mRNA expression (de Sousa Abreu, Penalva et al. 2009). Also, expression of all six genes was higher in myometrium and UL with the minor allele compared to myometrium and UL with the major allele (Figure 2-11b, c). Similarly, it is unclear if this difference is of relevance due to the fact that it is relatively small, is observed for every gene analyzed, and the variation between samples is relatively large.

DISCUSSION

The FGFF linkage study provides evidence for a genetic contribution to UL development as seen by two significant linkage peaks and several other suggestive peaks. Although candidate genes have yet to be identified under five peaks, one gene prominently recognized in UL biology, *HMGA2*, is located under the peak on 12q14. Translocations involving *HMGA2* are frequently found in UL tumors (Schoenberg Fejzo, Ashar et al. 1996). Additionally, a recent study using the FGFF population found a significant association between UL status and a variant in the 5' UTR of *HMGA2* (Hodge, K et al. 2009). The discovery of several linkage peaks illustrates the genetic heterogeneity of these tumors and is consistent with the hypothesis that a variety of genes and pathways participate in UL development. Additionally, a genome-wide association study in a Japanese cohort found three loci significantly associated with UL diagnosis: 10q24.33, 22q13.1, and 11p15.5 (Cha, Takahashi et al.). These loci are not associated with UL in our cohorts of white women, and our locus at 17q25.3 was not identified in the Japanese study supporting genetic heterogeneity in UL predisposition between ethnic groups.

Association analyses using the WGHS and Australian cohorts may have identified more precisely the locus underlying the linkage signal on chromosome 17. Six SNPs on chromosome 17 were in the top associated SNPs in the WGHS analysis and proved significantly associated with UL status after meta-analysis of the WGHS and Australian cohorts. These SNPs are in LD across *FASN*, *CCDC57* and *SLC16A3*. At this time it cannot be determined if any of these markers are the causal SNP associated with UL

affection status or if they are in LD with the causal SNP. SLC16A3 is a member of the proton-linked monocarboxylate transporter family that facilitates transport of substrates across the plasma membrane. CCDC57 contains a coiled coil domain and may function by binding DNA. SLC16A3 and CCDC57 are minimally characterized, making it difficult to conjecture about their possible involvement in UL development. Neither of these genes nor proteins has been indicated in disease and our qPCR studies did not reveal any unusual mRNA expression in UL. Conversely, we found a large increase in FAS expression in UL compared to matched myometrial tissue. FAS has been extensively characterized and is the enzyme responsible for *de novo* fatty acid synthesis. It is most highly expressed in hormone-sensitive cells (Kusakabe, Maeda et al. 2000) and has been found to be regulated at both transcriptional and post-transcriptional levels. Sterol regulatory element binding transcription factor 1 (SREBP-1) is the primary transcription factor of *FASN* and is activated downstream of growth factor and hormone receptors (Eberle, Hegarty et al. 2004). FAS protein is stabilized by USP2a, an isopeptidase that deubiquitinates and prevents protein degradation. USP2a is overexpressed in some prostate tumors with high FAS expression (Graner, Tang et al. 2004) and could explain discordant levels of *FASN* mRNA and FAS protein levels in our UL samples.

Upregulation of FAS has been discovered in many cancers including prostate, breast, and colon (Alo, Visca et al. 1999; Rossi, Graner et al. 2003; Ogino, Nosho et al. 2008). In some cancers, upregulation of FAS is correlated with poor prognosis, cancer progression or specific tumor types (Liu, Liu et al.). Inhibitors of FAS lead to growth arrest and apoptosis in cancer cell lines with relatively minor effects in corresponding

normal cells (Pizer, Jackisch et al. 1996; Pizer, Chrest et al. 1998). Many studies have found a connection between FAS and the PI3K/Akt signaling pathway, one of the most frequently disregulated pathways in human cancers (Van de Sande, De Schrijver et al. 2002; Porstmann, Griffiths et al. 2005). Inhibitors of FAS used in breast cancer animal models and several xenograft models have resulted in delayed development and slowed progression of tumors (Alli, Pinn et al. 2005; Lupu and Menendez 2006). Knocking down *FASN* mRNA causes a host of changes in gene expression and protein activity in the cell (Knowles and Smith 2007) making it clear from these and many more studies that the role of FAS in neoplasia is much more complicated and probably more important than simply providing fatty acids. Although it remains to be known how the minor allele of the candidate LD block influences *FASN* and UL development, overexpression of FAS in these UL samples and the overwhelming evidence that FAS is a metabolic oncogene makes it a compelling candidate gene for UL development for which clinical trials with inhibitors might be warranted.

MATERIALS AND METHODS

Finding Genes for Fibroids Linkage Analysis

Sister pairs affected with UL were recruited for the "Finding Genes for Fibroids" (FGFF) study. Approximately 385 sister pairs were consented for this project. Both sisters have medical record confirmed UL, provide a blood sample and complete a questionnaire on clinical, reproductive, sexual, and family history relating to UL. Other family members of the sisters also contributed samples and completed questionnaires. Study participants were recruited under an IRB protocol approved by the Partners HealthCare System Human Research Committee. DNA was isolated using a Puregene Blood Kit (Gentra, Minneapolis, MN) and the DNA, pedigree information, and UL affection status were provided to the genotyping core at The Center for Inherited Disease Research (CIDR) at Johns Hopkins University. A whole genome SNP linkage scan was performed using Illumina's Human Linkage-12 Genotyping BeadChip (San Diego, CA). Two families with multiple Mendelian inconsistencies were excluded in addition to 14 SNP markers due to low quality genotype calls.

Linkage analysis was performed with Genehunter software using samples from self-reported white sister pairs and family members, which comprised 261 families with a total of 1103 individuals. The minor allele frequency (MAF) of each SNP was calculated using the genotype information from one sister in each family. SNPs were pruned using PLINK software (Purcell, Neale et al. 2007) based on an r^2 of 0.2 resulting in a total of 4,196 SNPs for the final analysis. Sib-pair analysis was carried out with UL status as the phenotype and significant linkage with UL was defined as a LOD score greater than 3.6.

Despite particular interest in recruitment of black women, the FGFF study has not yet reached an appropriate number of sister pairs for this important analysis.

Women's Genome Health Study Association Study

The Women's Genome Health Study (WGHS) (Ridker, Chasman et al. 2008) is a prospective cohort of female North American health care professionals representing participants in the Women's Health Study (WHS) who provided a blood sample at baseline and consent for blood-based analyses. Participants in the WHS were 45 years of age or older at enrollment and free of cardiovascular disease, cancer or other major chronic illness. Additional information related to health and lifestyle were collected by questionnaires throughout the WHS trial and continuing observational follow-up. WHS participants were asked if they had ever been diagnosed with UL, their age at diagnosis, whether their mother or sister had ever been diagnosed with UL and their history of hysterectomy.

Genotyping in the WGHS sample was performed using the HumanHap300 Duo ''+'' chips or the combination of the HumanHap300 Duo and iSelect chips (Illumina, San Diego, CA) with the Infinium II protocol. In either case, the custom SNP content was the same; these custom SNPs were chosen without regard to minor allele frequency (MAF) to saturate candidate genes for cardiovascular disease as well as to increase coverage of SNPs with known or suspected biological function, *e.g*. disease association, non-synonymous changes, substitutions at splice sites, etc. For quality control, all samples were required to have successful genotyping using the BeadStudio v. 3.3 software (Illumina) for at least 98% of the SNPs. A subset of 23,294 individuals were identified with self-reported European ancestry, verified on the basis of multidimensional scaling

analysis of identity by state using 1,443 ancestry informative markers in PLINK v. 1.06 (Purcell, Neale et al. 2007). The final dataset of these individuals included SNPs with MAF >1%, successful genotyping in 90% of subjects, and deviations from Hardy-Weinberg equilibrium not exceeding $p=10^{-6}$. Among the final 23,294 individuals of verified European ancestry, genotypes for a total of 2,608,509 SNPs were imputed from the experimental genotypes and LD relationships implicit in the HapMap r. 22 CEU samples. Imputed SNPs were used to define the region of LD surrounding association signals found with genotyped SNPs.

UL status, age at diagnosis, mother or sister UL status, and history of hysterectomy were ascertained by recall in the 2009 WGHS questionnaire. UL cases and controls from WGHS were stratified based on these four variables in an attempt to identify women most and least likely to have a genetic basis for UL. Any participant who answered "not sure" for UL status or mother/sister UL status was excluded from the analysis. Participants were also excluded who reported an age of UL diagnosis at less than 20 years or at greater than 70 years. Cases included women who answered "yes" for UL status, "yes" for their mother or sister UL status, and either had an age of diagnosis under 40 years or had a hysterectomy. Controls included women who answered "no" to UL status, mother/sister UL status and who had not had a hysterectomy. Women not qualifying for either of these groups were excluded from the analysis. Following stratification there were 746 cases and 4,487 controls. Association analysis was performed on the set of 339,187 genotyped SNPs in PLINK using the standard case/control test and p-values less than 5×10^{-8} were considered significant.

Australian Cohort Association Study

Individuals comprising the cohort from the Queensland Institute of Medical Research (QIMR) were women who had been genotyped previously on Illumina's 317K, 370K or 610K SNP platforms as part of a larger collection of genome-wide association studies conducted at QIMR (Painter, Anderson et al.; Medland, Nyholt et al. 2009). Case samples (n = 484) were selected from amongst women originally recruited into a study of genetic factors underlying endometriosis (Treloar, Wicks et al. 2005) and a twin study of gynecological health (Treloar, Do et al. 1999). For both studies, women had completed questionnaires on various aspects of reproductive health, and cases had answered 'yes' to the 'uterine fibroids' option of the question "Have you ever had any of the following conditions?" Controls (n = 610) were taken from amongst twin pairs from the gynecological health study where both sisters had answered 'no' to the question on uterine fibroids (one sample per twin pair). A standard case/control association analysis was performed on the set of 269,629 SNPs genotyped in common between all samples and passing all QC metrics in PLINK (Purcell, Neale et al. 2007). Approval for the studies was granted by the Human Research Ethics Committee at QIMR and the Australian Twin Registry. All gene annotations and base pair positions are derived from the human genome sequence hg18 (NCBI build 36.1).

Demographics

Demographics of the FGFF, WGHS and Australian cohorts were considered, focusing on variables most relevant to UL diagnosis. In the WGHS and Australian cohorts, cases have a slightly higher BMI and have decreased stature compared to controls although the difference is not significant in the WGHS population (defined as a

Table 2-5. Demographics of the FGFF, WGHS and Australian cohorts

N		WGHS all (5233)	WGHS cases (746)	WGHS controls (4487)	P-value	Australian all (1094)	Australian cases (484)	Australian controls (610)	P-value	FGFF sisters (522)
Age in years (yrs)	Yrs Median (IQR)	52.9 (48.9-59.0)	52.3 (48.8-56.9)	52.5 (48.7-58.0)	0.0038	41.0 (34.0-50.0)	44.0 (37.0-51.0)	39.0 (34.0-48.0)	0.0001	47.0 (41.0-53.0)
BMI	kg/m^2 Median (IQR)	24.8 (22.3-28.3)	25.0 (22.6-28.2)	24.2 (22.0-27.4)	0.1595	23.6 (21.5-27.1)	24.8 (22.3-28.3)*	23.4 (21.3-26.5)	0.0019	24.1 (22.0-27.4)**
Height	Inches Median (IQR)	65.0 (63.0-66.0)	64.0 (63.0-66.0)	65.0 (63.0-66.0)	0.0528	63.8 (61.8-65.4)	63.4 (61.8-65.4)*	63.8 (61.8-65.7)	0.0195	65.0 (63.0-66.75)
Age at menarche	Yr Median (IQR)	12.0 (12.0-13.0)	12.0 (11.0-13.0)	12.0 (12.0-13.0)	2.82E-05	13.0 (12.0-14.0)	13.0 (12.0-14.0)	13.0 (12.0-14.0)	0.0069	12.0 (12.0-13.0)
Hysterectomy history	Hysterectomy N (%)					312 (28.5%)	263 (56.1%)	49 (8.3%)	3.77E-28	165 (30.1%)

*BMI and height were only available for 175 Australian cases.

**BMI is only available for 491 FGFF sisters.

P >0.05) (Table 2-5). Previous studies have also reported a correlation with UL diagnosis and a higher BMI (Kharazmi, Fallah et al. 2007; Takeda, Sakata et al. 2008). WGHS and Australian cases also have a younger age at menarche which could be expected as UL are hormone dependent neoplasms and a younger age at menarche is associated with longer reproductive years and additional years of hormone exposure. Lastly, although Australian cases were not selected based on a history of hysterectomy, UL cases have a very significant increase in hysterectomy prevalence reflecting the fact that UL are the leading cause for this surgery. Overall, demographics of the WGHS and Australian cohorts were unremarkable and confirmed previously reported UL associations.

FAS Immunohistochemistry

Fatty acid synthase (FAS) expression in UL and myometrium tissue was analyzed by immunohistochemistry of tissue microarrays (TMAs) consisting of paraffin embedded formalin fixed tissue sections from 200 women, 36 myometrium samples and 337 UL samples. The TMA includes matched myometrium and UL samples from 33 women. Two TMAs were constructed using sections from different areas of the same tissue. DNA corresponding to 106 tissue sections was isolated and used to genotype rs4247357 in *CCDC57*. Immunostaining was performed on both TMAs using the primary monoclonal antibody against FAS (Transduction Laboratories, Lexington, KY) at a 1:100 dilution and hematoxylin as a counterstain. Each core was evaluated for the ratio of stain to counterstain to account for variable cellularity in tissue sections. The average stain/counterstain ratio was determined and compared across myometrium and UL samples and also across samples with the major and minor allele of rs4247357.

mRNA Expression Analysis by Microarray

RNA was isolated from myometrium and 28 UL from 14 women. Gene expression levels were analyzed using the Affymetrix GeneChip system U133 plus 2.0. DNA was also extracted and used to genotype rs4247357 in *CCDC57*. Gene expression levels were compared between women with the major and minor allele of rs4247357.

Targeted mRNA Expression Analysis by qPCR

RNA was isolated from myometrium and 20 UL from 12 women, six with the major allele of rs4247357 and six with the minor allele. cDNA was synthesized and used in qPCR assessments of *FASN*, *CCDC57*, *SLC16A3*, *DUS1L*, *CSNK1D*, and *NARF*. Gene products were normalized to the internal reference gene, *GAPDH*, and normalized ratios were compared across myometrium and UL samples and across women with the major and minor allele of rs4247357.

Acknowledgements

The authors thank all of the women and their families who participated in the FGFF, WGHS and Australian studies. The FGFF study is supported by NICHD grants HD046226 and HD060530 (to C.C.M.). The WGHS study is supported by NIH grants CA047988, HL043851, HL080467, and HL099355. The Australian cohort was supported by National Institutes of Health Grants (AA07535, AA07728, AA13320, AA13321, AA14041, AA11998, AA17688, DA012854, and DA019951); by Grants from the Australian National Health and Medical Research Council (339462, 389927, 389938, 442981, 443036 and 496739); by the 5th Framework Programme (FP-5) GenomEUtwin Project (QLG2-CT-2002-01254); by the Wellcome Trust (WT085235/Z/08/Z); and by funding provided by the Cooperative Research Centre for Discovery of Genes for Common Human Diseases (CRC), Cerylid Biosciences (Melbourne). GWM, SEM and DRN are supported by the National Health and Medical Research Council Fellowship Scheme, and KTZ by a Wellcome Trust Research Career Development Fellowship. Genotyping services for FGFF were provided by the Center for Inherited Disease Research (CIDR). CIDR is funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. WGHS genotyping was provided by Amgen, Inc

REFERENCES

Alli, P. M., M. L. Pinn, et al. (2005). "Fatty acid synthase inhibitors are chemopreventive for mammary cancer in neu-N transgenic mice." Oncogene **24**(1): 39-46.

Alo, P. L., P. Visca, et al. (1999). "Fatty acid synthase (FAS) predictive strength in poorly differentiated early breast carcinomas." Tumori **85**(1): 35-40.

Altshuler, D., M. J. Daly, et al. (2008). "Genetic mapping in human disease." Science **322**(5903): 881-8.

Ashar, H. R., M. S. Fejzo, et al. (1995). "Disruption of the architectural factor HMGI-C: DNA-binding AT hook motifs fused in lipomas to distinct transcriptional regulatory domains." Cell **82**(1): 57-65.

Battista, S., V. Fidanza, et al. (1999). "The expression of a truncated HMGI-C gene induces gigantism associated with lipomatosis." Cancer Res **59**(19): 4793-7.

Borrmann, L., B. Seebeck, et al. (2003). "Human HMGA2 promoter is coregulated by a polymorphic dinucleotide (TC)-repeat." Oncogene **22**(5): 756-60.

Brosens, I., J. Deprest, et al. (1998). "Clinical significance of cytogenetic abnormalities in uterine myomas." Fertil Steril **69**(2): 232-5.

Buttram, V. C., Jr. and R. C. Reiter (1981). "Uterine leiomyomata: etiology, symptomatology, and management." Fertil Steril **36**(4): 433-45.

Cha, P. C., A. Takahashi, et al. "A genome-wide association study identifies three loci associated with susceptibility to uterine fibroids." Nat Genet **43**(5): 447-50.

Christacos, N. C., B. J. Quade, et al. (2006). "Uterine leiomyomata with deletions of Ip represent a distinct cytogenetic subgroup associated with unusual histologic features." Genes Chromosomes Cancer **45**(3): 304-12.

Cramer, S. F. and A. Patel (1990). "The frequency of uterine leiomyomas." Am J Clin Pathol **94**(4): 435-8.

de Sousa Abreu, R., L. O. Penalva, et al. (2009). "Global signatures of protein and mRNA expression levels." Mol Biosyst **5**(12): 1512-26.

Eberle, D., B. Hegarty, et al. (2004). "SREBP transcription factors: master regulators of lipid homeostasis." Biochimie **86**(11): 839-48.

Everitt, J. I., D. C. Wolf, et al. (1995). "Rodent model of reproductive tract leiomyomata. Clinical and pathological features." Am J Pathol **146**(6): 1556-67.

Fedele, L., F. Parazzini, et al. (1995). "Recurrence of fibroids after myomectomy: a transvaginal ultrasonographic study." Hum Reprod **10**(7): 1795-6.

Flynn, M., M. Jamison, et al. (2006). "Health care resource use for uterine fibroid tumors in the United States." Am J Obstet Gynecol **195**(4): 955-64.

Gattas, G. J., B. J. Quade, et al. (1999). "HMGIC expression in human adult and fetal tissues and in uterine leiomyomata." Genes Chromosomes Cancer **25**(4): 316-22.

Graner, E., D. Tang, et al. (2004). "The isopeptidase USP2a regulates the stability of fatty acid synthase in prostate cancer." Cancer Cell **5**(3): 253-61.

Gross, K. L. and C. C. Morton (2001). "Genetics and the development of fibroids." Clin Obstet Gynecol **44**(2): 335-49.

Gross, K. L., C. I. Panhuysen, et al. (2004). "Involvement of fumarate hydratase in nonsyndromic uterine leiomyomas: genetic linkage analysis and FISH studies." Genes Chromosomes Cancer **41**(3): 183-90.

Grosschedl, R., K. Giese, et al. (1994). "HMG domain proteins: architectural elements in the assembly of nucleoprotein structures." Trends Genet **10**(3): 94-100.

Hindley, J., W. M. Gedroyc, et al. (2004). "MRI guidance of focused ultrasound therapy of uterine fibroids: early results." AJR Am J Roentgenol **183**(6): 1713-9.

Hodge, J. C., T. C. K, et al. (2009). "Uterine leiomyomata and decreased height: a common HMGA2 predisposition allele." Hum Genet **125**(3): 257-63.

Hurst, B. S., D. J. Stackhouse, et al. (2000). "Uterine artery embolization for symptomatic uterine myomas." Fertil Steril **74**(5): 855-69.

Huyck, K. L., C. I. Panhuysen, et al. (2008). "The impact of race as a risk factor for symptom severity and age at diagnosis of uterine leiomyomata among affected sisters." Am J Obstet Gynecol **198**(2): 168 e1-9.

Kharazmi, E., M. Fallah, et al. (2007). "Cardiovascular diseases attributable to hysterectomy: a population-based study." Acta Obstet Gynecol Scand **86**(12): 1476-83.

Kjerulff, K. H., P. Langenberg, et al. (1996). "Uterine leiomyomas. Racial differences in severity, symptoms and age at diagnosis." J Reprod Med **41**(7): 483-90.

Knowles, L. M. and J. W. Smith (2007). "Genome-wide changes accompanying knockdown of fatty acid synthase in breast cancer." BMC Genomics **8**: 168.

Kusakabe, T., M. Maeda, et al. (2000). "Fatty acid synthase is expressed mainly in adult hormone-sensitive cells or cells with high lipid metabolism and in proliferating fetal cells." J Histochem Cytochem **48**(5): 613-22.

Lepine, L. A., S. D. Hillis, et al. (1997). "Hysterectomy surveillance--United States, 1980-1993." MMWR CDC Surveill Summ **46**(4): 1-15.

Lettre, G., A. U. Jackson, et al. (2008). "Identification of ten loci associated with height highlights new biological pathways in human growth." Nat Genet **40**(5): 584-91.

Ligon, A. H., S. D. Moore, et al. (2005). "Constitutional rearrangement of the architectural factor HMGA2: a novel human phenotype including overgrowth and lipomas." Am J Hum Genet **76**(2): 340-8.

Lingaas, F., K. E. Comstock, et al. (2003). "A mutation in the canine BHD gene is associated with hereditary multifocal renal cystadenocarcinoma and nodular dermatofibrosis in the German Shepherd dog." Hum Mol Genet **12**(23): 3043-53.

Liu, H., J. Y. Liu, et al. "Biochemistry, molecular biology, and pharmacology of fatty acid synthase, an emerging therapeutic target and diagnosis/prognosis marker." Int J Biochem Mol Biol **1**(1): 69-89.

Luoto, R., J. Kaprio, et al. (2000). "Heritability and risk factors of uterine fibroids--the Finnish Twin Cohort study." Maturitas **37**(1): 15-26.

Lupu, R. and J. A. Menendez (2006). "Pharmacological inhibitors of Fatty Acid Synthase (FASN)--catalyzed endogenous fatty acid biogenesis: a new family of anti-cancer agents?" Curr Pharm Biotechnol **7**(6): 483-93.

Marshall, L. M., D. Spiegelman, et al. (1997). "Variation in the incidence of uterine leiomyoma among premenopausal women by age and race." Obstet Gynecol **90**(6): 967-73.

Matta, W. H., R. W. Shaw, et al. (1989). "Long-term follow-up of patients with uterine fibroids after treatment with the LHRH agonist buserelin." Br J Obstet Gynaecol **96**(2): 200-6.

Medland, S. E., D. R. Nyholt, et al. (2009). "Common variants in the trichohyalin gene are associated with straight hair in Europeans." Am J Hum Genet **85**(5): 750-5.

Menendez, J. A. and R. Lupu (2007). "Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis." Nat Rev Cancer **7**(10): 763-77.

Nowak, R. A. (2001). "Identification of new therapies for leiomyomas: what in vitro studies can tell us." Clin Obstet Gynecol **44**(2): 327-34.

Ogino, S., K. Nosho, et al. (2008). "Cohort study of fatty acid synthase expression and patient survival in colon cancer." J Clin Oncol **26**(35): 5713-20.

Painter, J. N., C. A. Anderson, et al. "Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis." Nat Genet **43**(1): 51-4.

Pizer, E. S., F. J. Chrest, et al. (1998). "Pharmacological inhibitors of mammalian fatty acid synthase suppress DNA replication and induce apoptosis in tumor cell lines." Cancer Res **58**(20): 4611-5.

Pizer, E. S., C. Jackisch, et al. (1996). "Inhibition of fatty acid synthesis induces programmed cell death in human breast cancer cells." Cancer Res **56**(12): 2745-7.

Porstmann, T., B. Griffiths, et al. (2005). "PKB/Akt induces transcription of enzymes involved in cholesterol and fatty acid biosynthesis via activation of SREBP." Oncogene **24**(43): 6465-81.

Purcell, S., B. Neale, et al. (2007). "PLINK: a tool set for whole-genome association and population-based linkage analyses." Am J Hum Genet **81**(3): 559-75.

Quade, B. J., T. Y. Wang, et al. (2004). "Molecular pathogenesis of uterine smooth muscle tumors from transcriptional profiling." Genes Chromosomes Cancer **40**(2): 97-108.

Rein, M. S., A. J. Friedman, et al. (1991). "Cytogenetic abnormalities in uterine leiomyomata." Obstet Gynecol **77**(6): 923-6.

Ridker, P. M., D. I. Chasman, et al. (2008). "Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy american women." Clin Chem **54**(2): 249-55.

Rossi, S., E. Graner, et al. (2003). "Fatty acid synthase expression defines distinct molecular signatures in prostate cancer." Mol Cancer Res **1**(10): 707-15.

Schoenberg Fejzo, M., H. R. Ashar, et al. (1996). "Translocation breakpoints upstream of the HMGIC gene in uterine leiomyomata suggest dysregulation of this gene by a mechanism different from that in lipomas." Genes Chromosomes Cancer **17**(1): 1-6.

Stewart, L., G. M. Glenn, et al. (2008). "Association of germline mutations in the fumarate hydratase gene and uterine fibroids in women with hereditary leiomyomatosis and renal cell cancer." Arch Dermatol **144**(12): 1584-92.

Takeda, T., M. Sakata, et al. (2008). "Relationship between metabolic syndrome and uterine leiomyomas: a case-control study." Gynecol Obstet Invest **66**(1): 14-7.

Toro, J. R., M. L. Nickerson, et al. (2003). "Mutations in the fumarate hydratase gene cause hereditary leiomyomatosis and renal cell cancer in families in North America." Am J Hum Genet **73**(1): 95-106.

Treloar, S. A., K. A. Do, et al. (1998). "Genetic influences on the age at menopause." Lancet **352**(9134): 1084-5.

Treloar, S. A., K. A. Do, et al. (1999). "Predictors of hysterectomy: an Australian study." Am J Obstet Gynecol **180**(4): 945-54.

Treloar, S. A., J. Wicks, et al. (2005). "Genomewide linkage study in 1,176 affected sister pair families identifies a significant susceptibility locus for endometriosis on chromosome 10q26." Am J Hum Genet **77**(3): 365-76.

Van de Sande, T., E. De Schrijver, et al. (2002). "Role of the phosphatidylinositol 3'-kinase/PTEN/Akt kinase pathway in the overexpression of fatty acid synthase in LNCaP prostate cancer cells." Cancer Res **62**(3): 642-6.

Vikhlyaeva, E. M., Z. S. Khodzhaeva, et al. (1995). "Familial predisposition to uterine leiomyomas." Int J Gynaecol Obstet **51**(2): 127-31.

Williams, A. J., W. L. Powell, et al. (1997). "HMGI(Y) expression in human uterine leiomyomata. Involvement of another high-mobility group architectural factor in a benign neoplasm." Am J Pathol **150**(3): 911-8.

Zhou, X., K. F. Benson, et al. (1995). "Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C." Nature **376**(6543): 771-4.

CHAPTER 3

Investigating the role of a 5' UTR TC repeat polymorphism in *HMGA2* expression

S.L. Eggert[1], J.C. Hodge[4], C.C. Morton[2,3,5]

1) Department of Genetics, Harvard Medical School, Boston, MA
2) Departments of Obstetrics, Gynecology and Reproductive Biology and
3) of Pathology, Brigham and Women's Hospital, Boston, MA
4) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN
5) Harvard Medical School, Boston, MA

SLE performed all experiments and analyses. JCH obtained and genotyped samples. CCM contributed to overall study design and supervised project.

ABSTRACT

Uterine leiomyomata, commonly known as fibroids, are the most prevalent tumor in women of reproductive age. Genes implicated in UL biology have been identified through cytogenetic studies of the tumors and candidate gene association studies. One such gene, *HMGA2*, is involved in the most common translocation found in UL, t(12;14)(q14-15;q23-24), which results in its overexpression. A candidate gene association study of 18 markers across *HMGA2* revealed a significant association of a TC repeat polymorphism, corresponding to 27 TC repeats, in the 5' UTR with UL diagnosis. One published report describing the *HMGA2* promoter found the TC repeat polymorphism to be crucial for promoter activation. We investigated the role of the TC27 repeat in promoter activation using luciferase assays in 293T, UL and myometrial cell lines. No difference in promoter activation was found across seven common TC repeat polymorphisms, including TC27, in 293T cells. *HMGA2* promoter activation was greatly impaired in UL and myometrial cells which resulted in weak luminescence and large variability. The variability in this experiment prevented us from making a conclusion about the role of the TC27 repeat in UL and myometrial cells. Future work will focus on developing an alternative method of transfection in UL and myometrial cells to generate more robust luminescence.

INTRODUCTION

UL are independent, clonal tumors of the uterine myometrium tissue. Approximately 40% of UL have nonrandom chromosomal aberrations including translocations and deletions. The most common translocation, t(12;14)(q14-15;q23-24), is found in 20% of karyotypically abnormal tumors. Positional cloning identified *HMGA2* at the breakpoint in 12q14.3 (Schoenberg Fejzo, Ashar et al. 1996). Translocations involving *HMGA2* lead to elevated expression of the gene (Gattas, Quade et al. 1999). HMGA2 is an architectural factor that influences the transcription of a variety of genes. It is involved in many cell processes including growth, proliferation, differentiation, and death so it is reasonable to conclude that overexpression of this protein plays a role in UL development (Grosschedl, Giese et al. 1994). Further, a boy carrying an inherited, truncated variant of HMGA2 displays an overgrowth phenotype including a large number of lipomas (Ligon, Moore et al. 2005) which is similar to the phenotype of mice that also carry a truncated form of the protein (Battista, Fidanza et al. 1999).

Familial and twin studies demonstrate a genetic component to UL susceptibility (Treloar, Do et al. 1999; Luoto, Kaprio et al. 2000). In order to evaluate a possible role for *HMGA2* in UL risk, a candidate gene association study was performed. Using 248 affected white sister-pairs, 18 markers across *HMGA2* were analyzed for association to UL status. One marker, in the 5' UTR, is a TC dinucleotide repeat with approximately 20 polymorphic human alleles ranging from 18 to 39 repeats. Only one marker proved to be significantly associated with UL status as defined by a p-value<0.05 (Table 3-1). The

Table 3-1. Association results of TC dinucleotide repeat in the 5' UTR of *HMGA2* with UL status. Taken from Hodge, K et al. 2009.

		UL Development	
Allele*	Allele Frequency (%)	Number of Families	P-value$^{\Psi}$
229	14.9	64	0.45672
235	14.0	49	0.80129
227	10.3	37	**0.0005**
231	10.3	35	0.44071
233	9.5	33	0.77955
225	9.5	39	0.07161
221	7.5	29	0.97524
237	5.9	24	0.87565

*Eleven alleles with ≤ 5% frequency not shown.

$^{\Psi}$Minimal p test = 0.00049; P-values ranged from 0.17 to 1.0.

candidate marker corresponds to 27 TC repeats (p-value=0.00049) (Hodge, K et al. 2009). UL from women with the TC27 repeat have slightly higher *HMGA2* expression (median = 5.39; interquartile range = [1.57, 8.58]; n=11) than UL from non-TC27 women (median 2.15; interquartile range = [0.08, 5.03]; n=47) which supports a role for the TC27 repeat in *HMGA2* expression (Figure 3-1) (Hodge, K et al. 2009).

The *HMGA2* promoter's influence on *HMGA2* expression was investigated in a study using luciferase assays. Fragments of the promoter lacking the TC repeat polymorphism exhibited significantly decreased promoter activation. Another assay evaluated three contructs with the full *HMGA2* promoter and TC repeat polymorphisms, corresponding to 1, 22, and 36 TC repeats. Promoter activation increased with increasing TC repeat number (Borrmann, Seebeck et al. 2003). These studies provide evidence for a role of the TC repeat polymorphism in *HMGA2* expression. However, this experiment used TC repeat polymorphisms not commonly found in the human population (Figure 3-2) and did not evaluate TC27. Additional luciferase assays are needed to explore the possible role of the TC27 repeat in *HMGA2* expression and UL development.

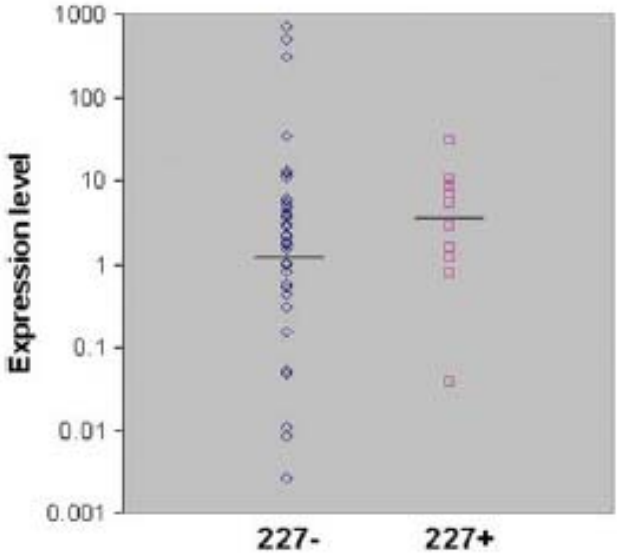


Figure 3-1. Real-time PCR expression results of *HMGA2* from UL of white women with and without the TC27 repeat polymorphism. Trend indicates higher *HMGA2* expression in women with TC27 but the results are not statistically significant. Taken from Hodge, K et al. 2009.

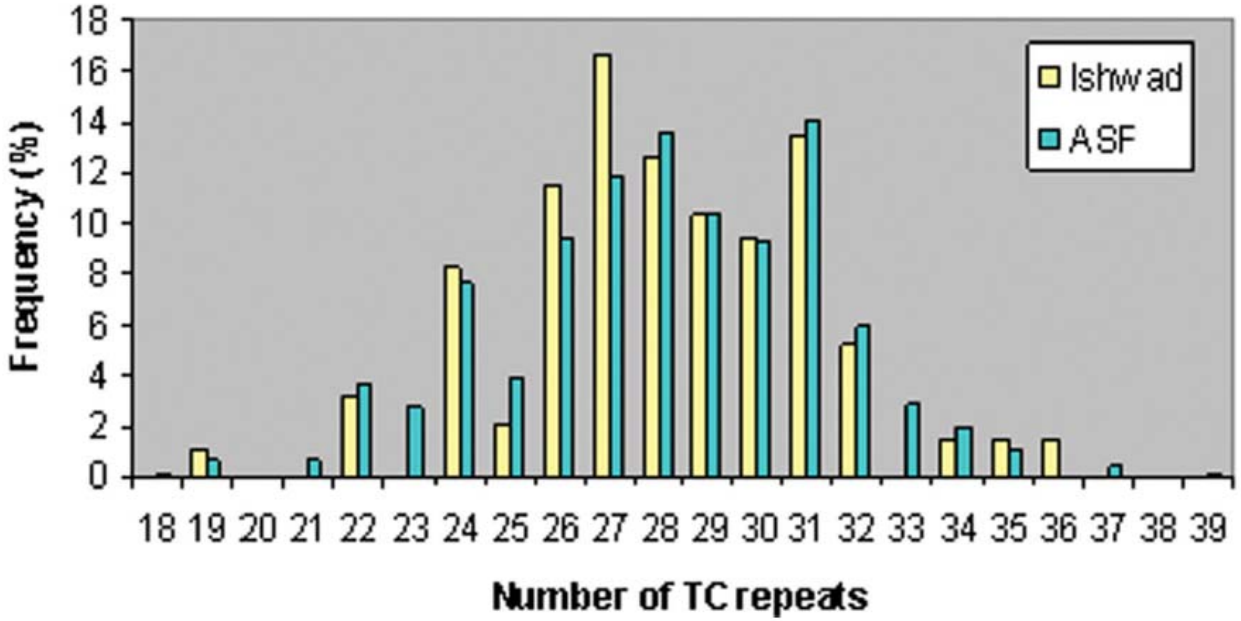


Figure 3-2. Frequency of TC repeat polymorphisms in the affected sister-pair cohort (ASF) and a previously published U.S. cohort (Ishwad et al. 1997). Taken from Hodge, K et al. 2009.

RESULTS

HMGA2 promoter activation in 293T cells

A luciferase assay was performed to correlate TC repeat number to *HMGA2* promoter activation. Three experiments were performed, each in triplicate, with the 293T cell line using ten constructs of different TC repeat number, a no DNA control, an empty vector control and a positive control (Figure 3-3). The empty vector control displayed very low promoter activation while the positive control exhibited significant activation. The *HMGA2* promoter constructs exhibited promoter activation only slightly lower than the positive control. These findings were consistent in replicates and across experiments. No significant difference in promoter activation was observed between the ten TC repeat polymorphisms in the *HMGA2* promoter.

HMGA2 promoter activation in UL and myometrium cells

Because we are concerned with the TC27 repeat in UL, we repeated the luciferase assay in normal myometrium and UL (with a t(12;14)) cell lines. Five experiments were performed, each in triplicate, using three TC repeat constructs, a no DNA control, an empty vector control and a positive control in the myometrium cell line (UtSMC-hTERT) (Figure 3-4) and the UL cell line (UtLM-hTERT) (Figure 3-5). Like the 293T cell line assays, the empty vector control displayed very low promoter activation compared to the positive control and the ratio between these controls was relatively consistent in replicates and between experiments. However, unlike the 293T cell line assays, the TC repeat constructs show much lower promoter activation compared to the positive control and the ratio of the TC repeat constructs to the positive control is hugely variable

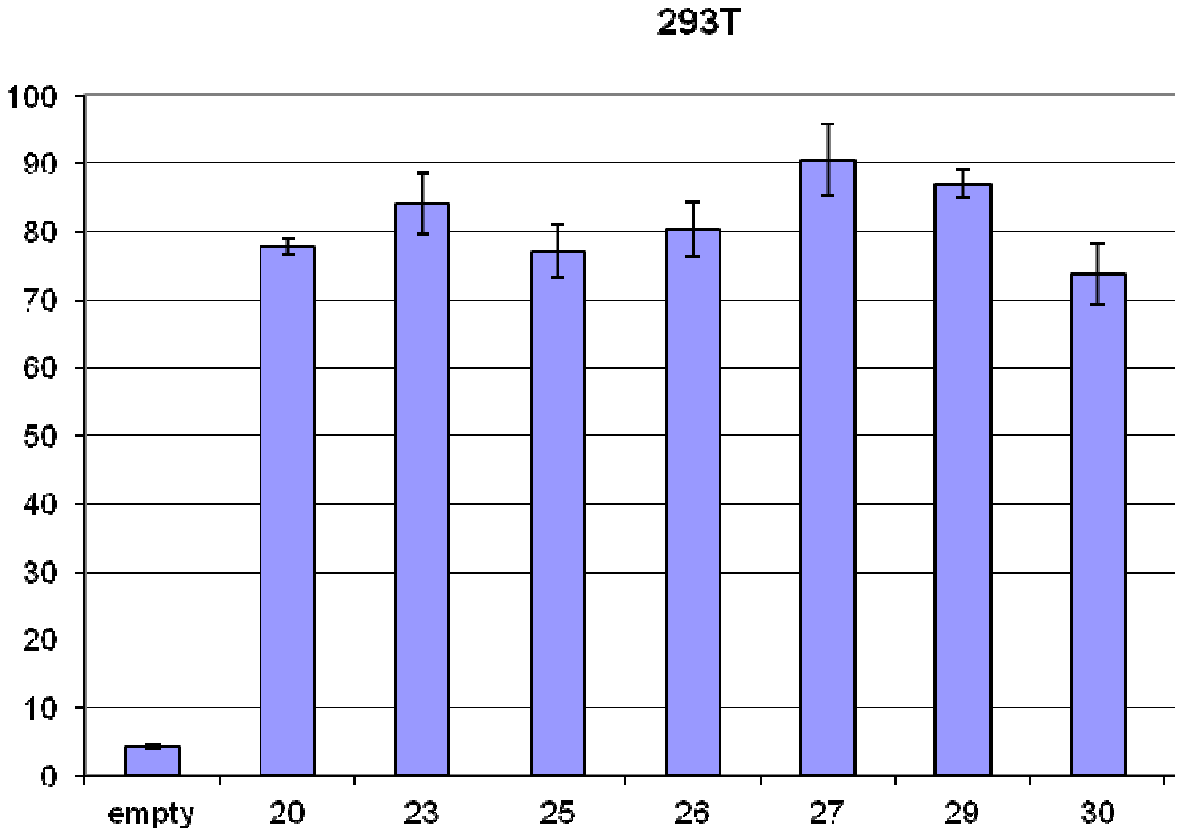


Figure 3-3. Luciferase assay results with 293T cells and several *HMGA2* promoter TC repeat polymorphisms. These data show no significant difference in promoter activation among TC repeats.

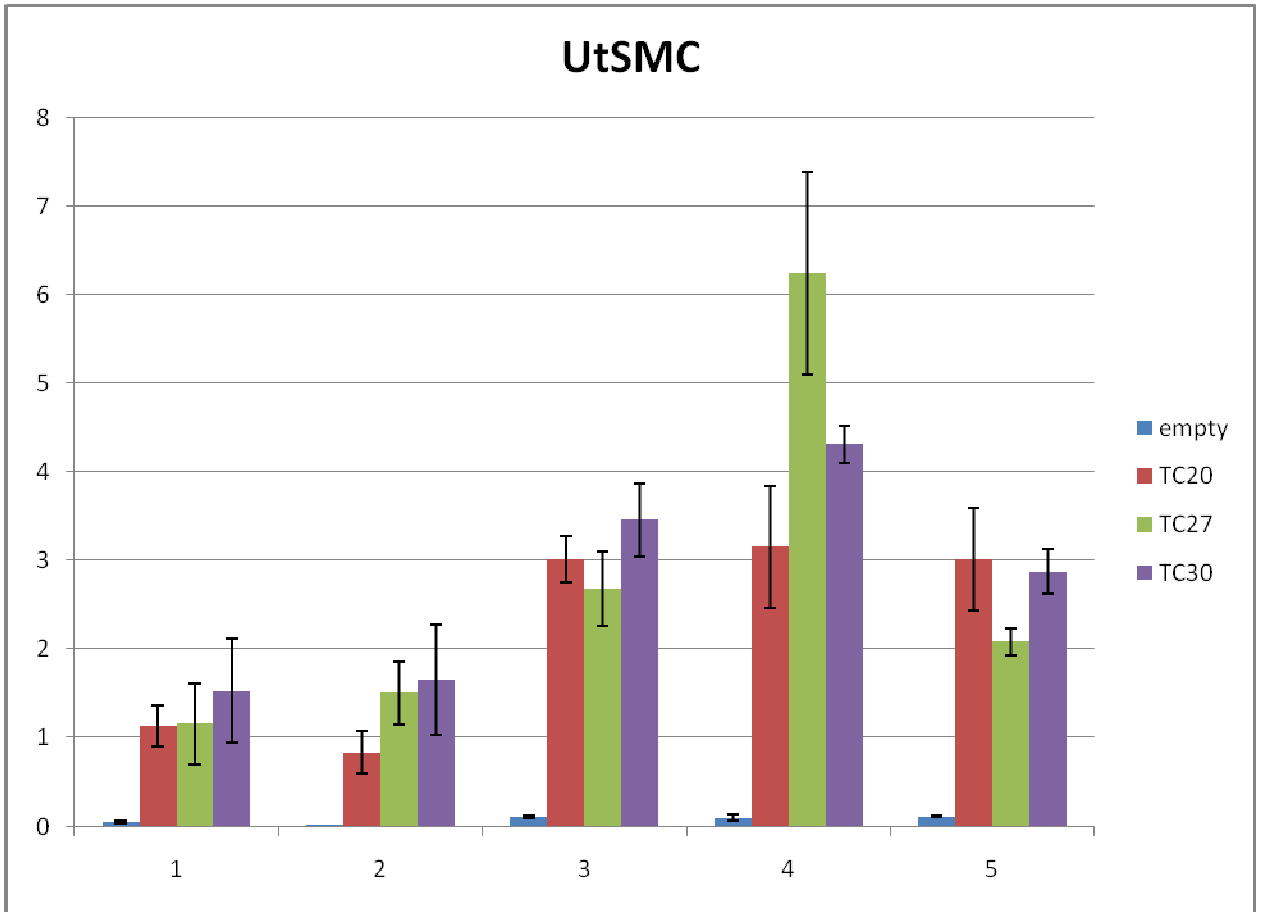


Figure 3-4. Luciferase assay results with UtSMC-hTERT cells and three *HMGA2* promoter TC repeat polymorphisms across five experiments. These data show hugely variable results in the TC repeat plasmids with relatively small error in the empty vector control.

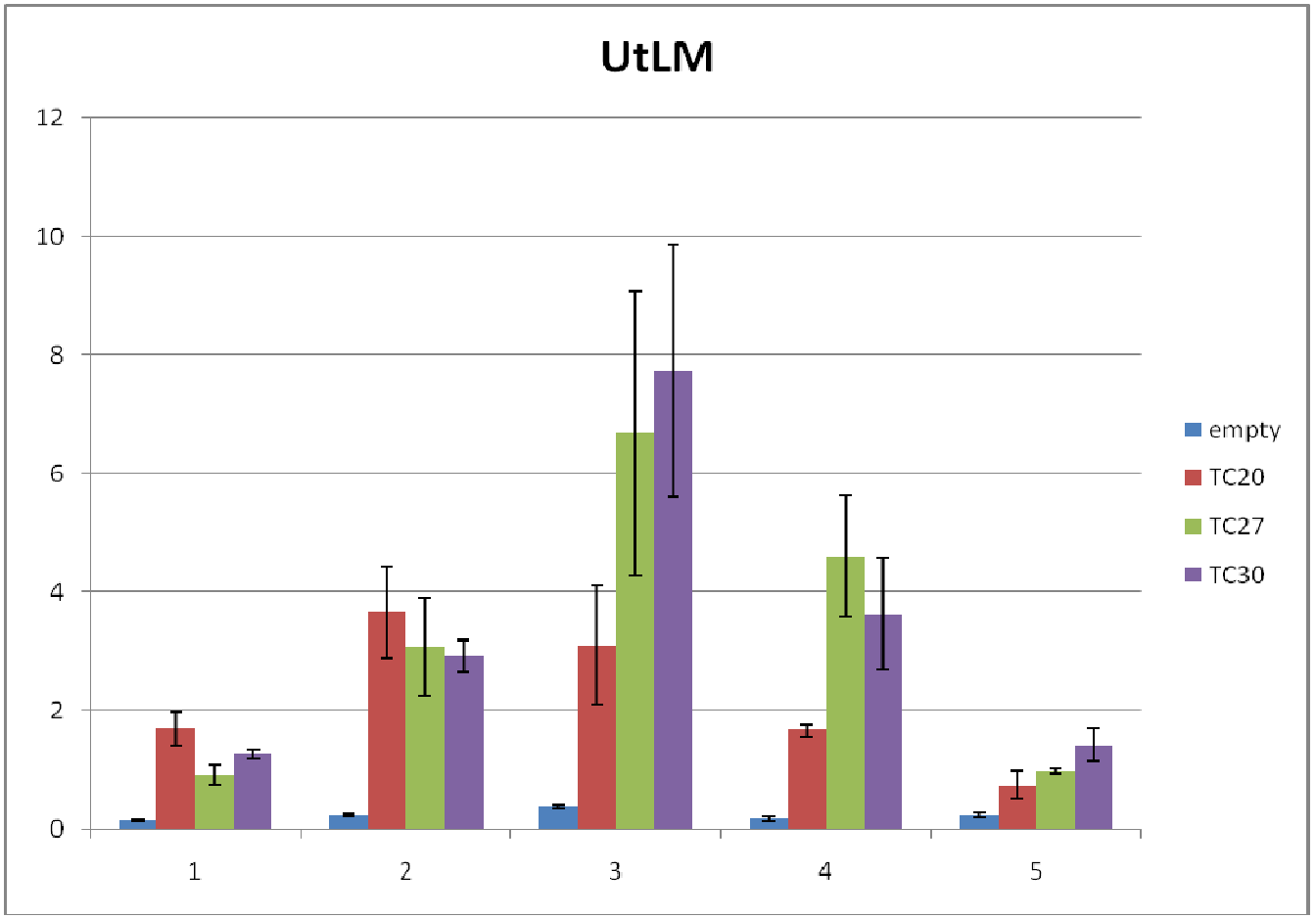


Figure 3-5. Luciferase assay results with UtLM-hTERT cells and three *HMGA2* promoter TC repeat polymorphisms across five experiments. These data show hugely variable results in the TC repeat plasmids with relatively small error in the empty vector control.

between replicates and across experiments. Each experiment in both cell lines also show hugely variable results and differing relationships between the TC repeat constructs. Although replicates of the positive control and the empty vector control are relatively consistent and show little error, the replicates of the TC repeat plasmids have large variance.

DISCUSSION

Overexpression of *HMGA2* is a common event in UL biology and multiple lines of evidence provide a link between inherited variants of *HMGA2* and overgrowth phenotypes. The TC27 repeat in the 5' UTR of *HMGA2* is significantly associated with UL diagnosis and may play a role in *HMGA2* expression. A study from the Borrmann group (Borrmann, Seebeck et al. 2003) showed an increase in *HMGA2* promoter activation with plasmids that contained larger TC repeats in a lipoma cell line. I found no significant difference in promoter activation across ten TC repeat polymorphisms after three luciferase assays in 293T kidney cells. Transcription regulation can vary by cell type and we were concerned that *HMGA2* promoter activation in 293T cells may not be relevant to UL biology so the experiment was repeated in myometrium and UL cell lines.

After five luciferase assays in myometrium and UL cell lines, I concluded that the results in these cell lines were too variable to determine whether or not there is a difference in promoter activation between the TC repeat constructs. The amount of luminescence produced from both the firefly and *Renilla* luciferase reporters were much higher in the 293T cells compared to the myometrium and UL cell lines, most likely indicating a higher transfection efficiency and possibly also elevated production of the reporters in the 293T cells. Additionally, the luciferase activity of the *HMGA2* TC repeat plasmids were measured at 70-90% compared to the positive control in the 293T cells and at less than 5% in the myometrium and UL cell lines suggesting *HMGA2* promoter activation is regulated differently in these cell lines. Although the total amount of luciferase activity was much lower in the myometrium and UL cell lines, the empty

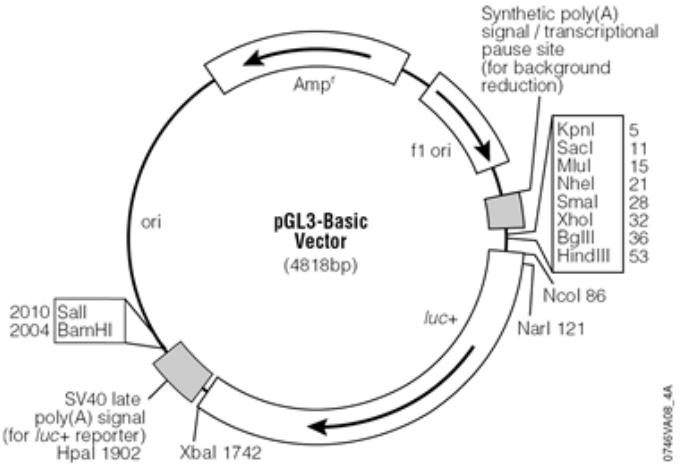
vector control was still relatively consistent while the TC repeat plasmids showed hugely variable results between experiments and also between replicates in the same experiment. This result provides evidence that there is an unknown variable in the luciferase assays with the *HMGA2* promoter plasmids in the myometrium and UL cell lines that has not been controlled in these experiments and that this variable was not a significant factor in the 293T cell assays.

Future work on this project will focus on developing an alternative method to investigate the role of the TC27 repeat on *HMGA2* expression in myometrium and UL cell lines. It is possible that a higher transfection efficiency will produce more robust results with less variation so an alternative method of transfection should be considered.

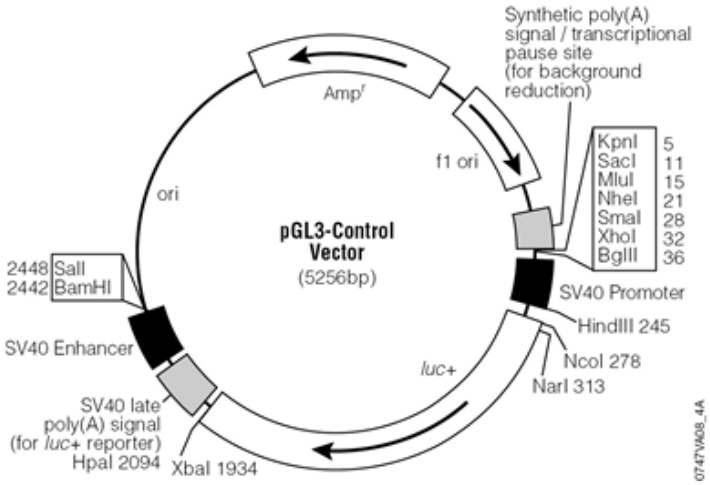
MATERIALS AND METHODS

Constructs were created using the pGL3 basic vector from Promega which contains the firefly luciferase gene with no promoter (Figure 3-6). Ten constructs were made containing the entire 5'UTR and promoter of *HMGA2* with different TC repeats corresponding to 20, 22, 23, 24, 25, 26, 27, 29, and 30 TC repeats. The pRL-TK vector from Promega was co-transfected with the experimental constructs as an internal control. The pRL-TK vector encodes the *Renilla* luciferase gene under the control of the herpes simplex virus thymidine kinase promoter and produces moderate levels of *Renilla* luciferase. A pGL3 basic vector with no promoter was used as a negative control and the pGL3 control vector, which contains the SV40 promoter and enhancer, was used as a positive control. Constructs were transiently transfected using the Lipofectamine 2000 kit into 293T cells, derived from human embryonic kidney cells, and into UtLM-hTERT, derived from a UL with a t(12;14), and UtSMC-hTERT, derived from normal myometrium cells. The luciferase assay was performed using Promega's Dual Luciferase Reporter Assay system and firefly and *Renilla* luminescence was measured 24 hours after transfection of the constructs. Experiments were performed in triplicate and results are reported as a ratio of firefly to *Renilla* luminescence and normalized to the positive control.

a



b



c

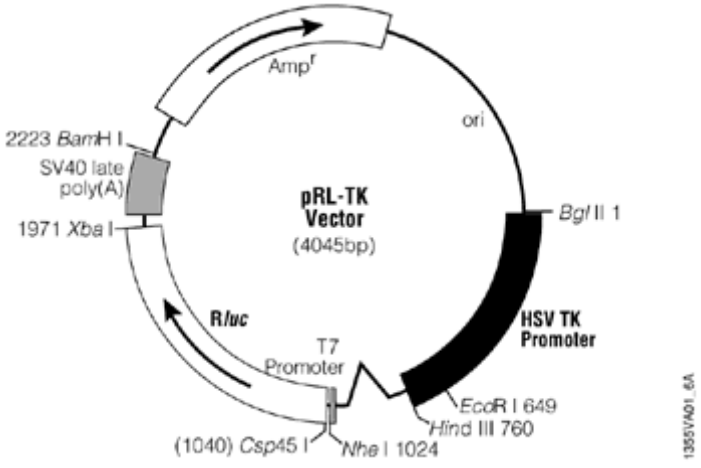


Figure 3-6. Schematics of a) the pGL3 basic vector used to create the experimental constructs, b) the pGL3-control vector used as a positive control, and c) the pRL-TK vector used as an internal control.

REFERENCES

Battista, S., V. Fidanza, et al. (1999). "The expression of a truncated HMGI-C gene induces gigantism associated with lipomatosis." Cancer Res **59**(19): 4793-7.

Borrmann, L., B. Seebeck, et al. (2003). "Human HMGA2 promoter is coregulated by a polymorphic dinucleotide (TC)-repeat." Oncogene **22**(5): 756-60.

Gattas, G. J., B. J. Quade, et al. (1999). "HMGIC expression in human adult and fetal tissues and in uterine leiomyomata." Genes Chromosomes Cancer **25**(4): 316-22.

Grosschedl, R., K. Giese, et al. (1994). "HMG domain proteins: architectural elements in the assembly of nucleoprotein structures." Trends Genet **10**(3): 94-100.

Hodge, J. C., T. C. K, et al. (2009). "Uterine leiomyomata and decreased height: a common HMGA2 predisposition allele." Hum Genet **125**(3): 257-63.

Ligon, A. H., S. D. Moore, et al. (2005). "Constitutional rearrangement of the architectural factor HMGA2: a novel human phenotype including overgrowth and lipomas." Am J Hum Genet **76**(2): 340-8.

Luoto, R., J. Kaprio, et al. (2000). "Heritability and risk factors of uterine fibroids--the Finnish Twin Cohort study." Maturitas **37**(1): 15-26.

Schoenberg Fejzo, M., H. R. Ashar, et al. (1996). "Translocation breakpoints upstream of the HMGIC gene in uterine leiomyomata suggest dysregulation of this gene by a mechanism different from that in lipomas." Genes Chromosomes Cancer **17**(1): 1-6.

Treloar, S. A., K. A. Do, et al. (1999). "Predictors of hysterectomy: an Australian study." Am J Obstet Gynecol **180**(4): 945-54.

CHAPTER 4:

CONCLUSION

In this thesis, I describe two genetic variants that associated with an increased risk of developing UL. We performed the first genome-wide scan for genetic risk variants for UL in white women and characterized a variant previously discovered in a candidate gene association study. Currently, hysterectomy is the most common treatment for UL. Understanding the pathways involved in UL biology is vital to develop novel, less invasive therapies. Further, associating genetic variants to particular UL features, such as size of tumors, severity of symptoms, and age at onset, will be useful for screening. Genetic screening for women with a family history of UL or who already exhibit symptoms could facilitate family planning and aid in determining optimal treatment options.

Using genome-wide linkage and association studies, we successfully identified one genetic variant significantly associated with UL diagnosis. Many limitations to these approaches can make it difficult to discover causal genes. The FGFF linkage results uncovered seven linkage peaks with significant and suggestive LOD scores but these peaks are very large and contain hundreds of genes. We have very appealing candidate genes for two of the peaks, *HMGA2* under 12q14 and *FASN* at 17q25. Variants in these genes have not been genotyped in the FGFF cohort at this time, therefore, we cannot explicitly demonstrate that these genes contribute to the linkage peaks. An LD block spanning *FASN* is significantly associated with UL cases after a meta-analysis of two GWASs. The quantile-quantile plot of the results supports this conclusion as the p-values of the SNPs in this LD block are much smaller than would be expected under the null hypothesis. The quantile-quantile plot reveals additional SNPs with smaller p-values than would be expected by chance; however, they do not meet the threshold for genome-

wide significance and cannot be considered significantly associated with UL. Certainly, many true associations have been overlooked because the GWASs did not have the power to detect them. Undoubtedly, a considerable amount of power was lost due to misclassification of controls and small sample size. Additional samples and medically confirmed controls will provide the power needed to detect additional variants associated with UL.

The top, significant SNP revealed in the meta analysis is located in a large LD block in 17q25.3 comprising numerous SNPs located in over approximately 150 kb. Many SNPs in the LD block are in very high LD with the top SNP complicating location of the causal variant. At this time it is difficult to conjecture about the role of the minor allele of this LD block in UL biology. Of the SNPs in the LD block described by HapMap data, there are no variants that are likely to affect FAS protein function. However, there is a high probability that some of the variants fall in regulatory regions. It is challenging to identify a causal regulatory variant which might include SNPs that affect gene transcription, RNA stability and protein translation. Further, the causal variant may not be described in the HapMap database and sequencing may be required to determine it. We found a pattern of increased FAS expression in myometrium and UL from women with the major allele of the top SNP compared to samples from women with the minor allele. This may indicate that the causal variant affects protein expression or stability, however, more work is needed to identify the causal variant and determine its effect on FAS.

Three genes are located in the candidate LD block, and our functional work and previous research supports a role for *FASN* in UL biology. Immunohistochemistry

results showing elevated FAS expression in UL is consistent with similar findings in other hormone-dependent tumors. Overexpression of FAS in breast and prostate cancer has been found to be important for tumor cell survival and for this reason we believe *FASN* is a very appealing candidate gene for UL. Inhibitors of FAS are attractive therapies for tumors overexpressing the protein because they seem to slow tumor progression with minor effects on normal cells which could mean less, more manageable side effects. In the case of UL, FAS inhibitors could be the first noninvasive option for tumor management. Future work will include investigating effects of FAS inhibitors on myometrial and UL cell lines.

The second part of my thesis work focused on the TC27 repeat polymorphism in *HMGA2* that was previously found to be associated with UL diagnosis. *HMGA2* has been associated with UL biology through chromosomal aberrations for many years and is known to be overexpressed in the tumors that harbor such aberrations. Inherited variants of *HMGA2* have also resulted in overgrowth phenotypes, making it an appealing candidate for risk of developing UL. The TC repeat in the 5'UTR of *HMGA2* was found to contribute to promoter activation in a luciferase assay in three cell lines and increasing TC repeat number resulted in increased promoter activation in a lipoma cell line. It is reasonable to conjecture that the TC27 repeat affects *HMGA2* expression which increases the risk of developing UL. Unfortunately, the method described in this thesis was not robust enough to evaluate the role of TC27 in the relevant myometrial and UL cell lines. Further work will include developing a more efficient method of transfection in the myometrial and UL cell lines in an effort to produce sufficient luminescence.

Unbiased, genome-wide studies have been very successful in uncovering genes associated with disorders that would not otherwise have been considered a candidate gene. Herein, I have described the first risk variant of UL identified in white women by a genome-wide approach. Undoubtedly, our linkage and association results harbor additional variants that contribute to risk for developing UL. Moving forward, whole genome sequencing of cases and controls will prove to be invaluable for uncovering the variants contributing to our UL linkage peaks. Sequence data will also be utilized to identify the causal variant in the candidate LD block at 17q25.3. Further, studies of additional cohorts of women with UL diagnosis and phenotypic information will increase the power of our meta analysis and reveal more variants with genome-wide significance for UL predisposition. Identifying risk variants for UL adds to our knowledge of the complex biology of these tumors and brings us closer to developing novel therapies.